UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Environmental & Geographical Science

Using an airborne Hyperspectral and LiDAR integrated sensor approach to spectrally discriminate and map savanna bush encroaching species in the Greater Kruger National Park Region

Laven Naidoo

Presented in fulfilment of the requirements for the MSc degree in Environmental and Geographical Science

Supervisors: Dr R Mathieu; Dr M Cho (CSIR)

Dr F Eckardt (UCT)

Date: June 2011

CSIR
our future through science

DEDICATIONS

I would like to dedicate this thesis to a dear friend, mentor and former supervisor, Mrs Shirley Butcher, who inspired me to pursue research in GIS and other geospatial scientific fields. Mrs Shirley Butcher sadly passed away at the beginning of this year from a long battle with cancer. You are sorely missed and will never be forgotten.

PLAGIARISM DECLARATION

1. *I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.*

2. *Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.*

3. *I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.*

4. *I declare that this thesis is my own work*

Table of Contents

Dedications 2

Plagiarism Declaration 3

Thesis Abstract 8

Chapter 1: General Introduction To Bush Encroachment, Species Mapping And CAO Hyperspectral & LiDAR Remote Sensing System 10
1.1 The Bush Encroachment Phenomenon 10
1.1.1 Impacts Of Bush Encroachment 11
1.1.2 History Of Bush Encroachment In The Kruger National Park Region 12
1.2 Remote Sensing As A Tool For Monitoring Bush Encroaching Species 13
1.2.1 Bush Encroaching Species Mapping Using Remote Sensing – A Shift In Perspective 13
1.2.2 The Challenges To Be Encountered In The Mapping Of Savanna Tree Species 15
1.2.3 Recent Developments In Remote Sensing 16
1.2.3.1 Biochemical Remote Sensing Of Vegetation 17
1.2.3.2 Structural Remote Sensing Of Vegetation 18
1.3 Carnegie Airborne Observatory (CAO) Integrated Sensor 19
1.4 Overall Study Aims, Objectives And Thesis Structure 20
1.4.1 Aim 20
1.4.2 Objective 20
1.4.3 Thesis Structure And Outline 21
1.4.3.1 Chapter 2 21
1.4.3.2 Chapter 3 21
1.4.3.3 Chapter 4 22
1.5 The General Study Area - The Greater Kruger National Park Study Region (L1-8) 22
1.5.1 Bushbuckridge Municipality District (BMD) 22
1.5.2 Sabi Sands Wildetuin 23
1.5.3 Skukuza Region, Kruger National Park 24
1.6 General Thesis Materials And Methods 26
1.6.1 CAO Data Description 26
1.6.2 Pre-Processing And Calibration Of CAO Data 26

Chapter 2: Spectral Discrimination Of Savanna Tree Species, With Emphasis On Bush Encroaching Species, Using Canopy Level Spectra From The CAO Hyperpsectral Data 27
Overview 27
2.1 Introduction 28
2.1.1 The Importance Of Tree Species Discrimination 28
2.1.2 Vegetation Indices And Statistical Measures 28
2.1.3 Spectral Similarity Measures 29
2.1.4 Chapter Aim And Research Questions 32
2.2 Methodology 33
2.2.1 April 2009 Field Campaign 33

2.2.2 Spectral Data Pre-Processing 34
2.2.3 Spectral Similarity Measure Formulae 35
2.2.4 Spectral Scenarios 36
2.2.5 Validation 37
2.3 Results 37
2.3.1 SAM Results 45
2.3.2 SID Results 46
2.3.3 SID-SAM Results 46
2.3.4 PSD Results 47
2.4 Discussion 49
2.5 Conclusions 51
2.6 Recommendations For Species Mapping Procedure 52

Chapter 3: The Mapping Of Savanna Bush Encroaching Species Using Multiple Endmember Spectral Angle Mapper In A Novel Spectral And Structural Decision Tree Approach 53
Overview 53
3.1 Introduction 54
3.1.1 Why Map Savanna Tree Species? 54
3.1.2 Available Techniques For Tree Species Mapping, Their Limitations And The Way Forward 55
3.1.3 Aims, Objectives And Research Questions 59
3.2 Methodology 60
3.2.1 May 2010 Field Campaign And Image Data Selection 61
3.2.2 Decision Tree Design Process 62
3.2.3 Strategies Taken To Improve Decision Tree Classification Accuracy 68
3.2.4 Bush Encroaching Tree Species Mapping 69
3.2.5 Summarised Methodology Schema 72
3.3 Results 73
3.3.1 Multivariable Histograms 73
3.3.2 Species Threshold Matrix 77
3.3.3 The Original Decision Tree Design Confusion Matrix 79
3.3.4 Evolution Of The Original Decision Tree Design (Figure 3.8) 82
3.3.5 Final Decision Tree Design Confusion Matrix 83
3.3.6 Bush Encroaching Species Maps Of The L456 Region 85
3.4 Discussion 90
3.5 Conclusions 98

Chapter 4: Thesis Synthesis And Recommendations 100
4.1 Study Aims And Objectives Revisited 100
4.1.1 Savanna Tree And Bush Encroaching Species Discrimination 100
4.1.2 Mapping Of Savanna Bush Encroaching Species 101
4.2 Study Recommendations 101
4.2.1 Recommendations For Chapter 2 101
4.2.2 Recommendations For Chapter 3 102

Acknowledgements 103

References 105

Appendices.....114
1a. CAO System Configuration.....114
1b. CAO Sensor System Specifications (Summarized).....115
2a. Spectral Similarity Measure Formulae.....115
2b. Plant Chemical Components And Their Associated Wavelengths.....116
2c. 49 Savanna Tree Species List And Accompanying Abbreviations.....116
2d. SAM Matlab Script.....118
2e. SID Matlab Script.....118
3a. 23 Savanna Tree Species Information: Codes Utilized In The Results Section And Attribute Information (Schmidt Et Al, 2007 & Shackelton Et Al, 2005)..120
3b. Total Number Of Sampled Pixels, Per Species, For The Training And Validation Datasets.....124
3c. Confusion Matrix – Minor Pruning Strategy.....125
3d. Confusion Matrix – Minor Pruning Strategy + Selected Training Endmembers.....126
3e. Confusion Matrix – Minor Pruning Strategy + Selected Significant Bands 127
3f. Confusion Matrix – Traditional Multiple Endmember Sam (No Decision Tree).....128

List of Figures

Figure 1.1: Map of the Greater Kruger National Park Study Area illustrating the different study regions (L1-L8) and their dominant geology .....**Error! Bookmark not defined.**5
Figure 2.1: Overall accumulated SAM values across all 14 spectral scenarios.....39
Figure 2.2: SAM values for different species across 3 dominant spectral scenarios...40
Figure 2.3: Overall accumulated SID values across all 14 spectral scenarios.....411
Figure 2.4: SID values for different species across 3 different spectral scenarios 422
Figure 2.5: Overall accumulated SID-SAM (tan) values across all 14 spectral scenarios.....433
Figure 2.6: SID-SAM (tan) values for different species across 3 different spectral scenarios.....444
Figure 2.7: Overall PSD results for all the species for SAM, SID & SID-SAM (tan) measures.....48
Figure 3.1: Pre-selected and sampled tree canopies of L456 which assisted in the development of the spectral and structural libraries…….62
Figure 3.2: *Sclerocarya birrea* reflectance spectra from Sabi Sands and Bushbuckridge…….64
Figure 3.3: *Combretum apiculatum* reflectance spectra from Sabi Sands and Bushbuckridge…….64
Figure 3.4: Original 9 End Node Decision Tree Design……..67
Figure 3.5: Methodology schema for decision tree construction, classification and post-classification…….72
Figure 3.6: Minimum tree height histogram of the 23 savanna tree species……74
Figure 3.7: Vegetation indices histograms of the 23 savanna tree species - (A) NDVI; (B) PRI; (C) CRI & (D) Red Edge NDVI……...75
Figure 3.8: Evolution of the original decision tree design……...82
Figure 3.9: Original L456 hyperspectral image with delineated areas of interest…...86

Figure 3.10: Species distribution map of the bush encroacher, *Acacia gerrardii*……87
Figure 3.11: Species distribution map of the bush encroacher, *Combretum apiculatum*……87
Figure 3.12: Species distribution map of the bush encroacher, *Dichrostachys cinerea*……..88
Figure 3.13: Species distribution map of the bush encroacher, *Terminalia sericea*…88
Figure 1A: The CAO integrated sensor platform……...115

List of Tables

Table 3.1: The four Vegetation Indices used and their associated formulae and references……..65
Table 3.2: Strategies undertaken to improve decision tree classification accuracy…..69
Table 3.3: Species spectral and structural parameter threshold matrix……78
Table 3.4: Original 9 End Node Decision Tree confusion matrix……80
Table 3.5: Final 2 End Node Decision Tree confusion matrix…….83

THESIS ABSTRACT

Bush encroachment is an environmental phenomenon which affects arid and semi-arid savanna rangelands across the world. Bush encroachment has numerous negative and positive impacts on these savanna ecosystems depending on the land use practices and associated rangeland management regimes. In an attempt to understand bush encroachment further, mapping of encroaching species, rather than traditional quantitative woodiness analyses, will provide a greater understanding of savanna ecosystem dynamics for management. Remote sensing, being the most appropriate tool for regional land cover monitoring and analyses, have traditionally lacked the necessary technology, both sensor and methods wise, to achieve accurate species level mapping in complex environments up until now. The introduction of the Carnegie Airborne Observatory (CAO) integrated hyperspectral and Light Detection And Ranging (LiDAR) sensor technology have provided simultaneous spectral and structural information to mitigate the challenges associated with tree species level mapping in savanna environments. This study aims to spectrally decompose and describe the regional patterning of key bush encroaching species in the Greater Kruger National Park savanna ecosystem of Mpumalanga, South Africa. Before mapping can be achieved, the spectral discrimination or separability between bush encroaching and other savanna tree species need to be explored and understood.

Three spectral similarity measures [Spectral Angle Mapper (SAM), Spectral Information Divergence (SID) & SID-SAM mixed measure] were applied across 14 different spectral scenarios to determine which spectral region was best suited for savanna tree species discrimination and which measure yielded the highest discriminatory results. It was found that SAM yielded the greatest probability of spectrally discriminating most of the savanna tree species, including bush encroaching species, under investigation with the bulk of the discriminating potential present in the entire transformed spectrum and the primary and secondary plant chemical band scenarios. For the mapping procedure, a novel decision tree multiple endmember SAM classification approach was proposed and tested against a traditional multiple endmember SAM. The final 2 end node decision tree design derived from minimum tree height thresholding, together with the use of spectrally significant bands, yielded the highest overall classification accuracy of 62.2% which

surpassed the traditional approach (56.7%). The individual bush encroaching species maps, produced from the final decision tree design, indicated that Dichrostachys cinerea was the most prolific encroacher species with a dominant presence across the study region. Combretum apiculatum was the least abundant but followed a distribution pattern similar to Dichrostachys cinerea. Terminalia sericea, a recognised seep-line species, displayed a distinct but strict southerly distribution in the study region which differed noticeably from the other species.

CHAPTER 1: GENERAL INTRODUCTION TO BUSH ENCROACHMENT, SPECIES MAPPING AND CAO HYPERSPECTRAL & LiDAR REMOTE SENSING SYSTEM

1.1 THE BUSH ENCROACHMENT PHENOMENON

Bush encroachment is described as a global phenomenon which has been observed on every continent where arid and semi-arid rangelands occur (Archer et al, 1995 cited in Hudak & Wessman, 1998; Oldeland et al, 2010). According to Ward (2005), bush encroachment is defined as the suppression of palatable grasses and herbs by encroaching woody species often unpalatable to domestic livestock. This phenomenon is a major concern in the protected and conserved savanna ecoregions of the world (e.g. Kruger National Park and Sabie Sands Wildetuin) as there is an erratic but stable equilibrium of co-existence between the dominance of trees (woody cover) and grasses. This equilibrium has shifted from a state of tree and grass co-existence to one which is slowly becoming dominated by trees and other woody vegetation (Dr E. February, personal communication – May 2007). In such protected areas, human impact and presence is strictly managed and restricted but the opposite can be said for the neighbouring communal rangelands (e.g. Bushbuckridge area). These communal areas have been exposed to over-utilisation (overgrazing and overharvesting of fuelwood) due to high human population pressures in confined areas that originated from the resettlement and segregation policies of the former Apartheid government (Wessels et al, 2004). It is thus paramount to assess the severity of bush encroachment in both protected and communal environmental settings as both human livelihoods, wild life and ecosystem processes would definitely be at stake. Bush encroachment is traditionally perceived negatively as a form of land degradation (Oldeland et al, 2010) but the impacts of this phenomenon could be viewed as positive and/or negative depending on the land management strategies and the livelihoods of the local populace.

1.1.1 IMPACTS OF BUSH ENCROACHMENT

Bush encroachment has been described to have numerous detrimental effects on the surrounding environments. According to Ward (2005), bush encroachment significantly reduces the carrying capacity of the land for livestock due to the replacement of palatable grasses with woody cover. As a result of this, livestock productivity declines which subsequently threatens local livelihoods and businesses of the farmers and surrounding communities (Hudak & Wessman, 2001). The sustainability of the pastoral subsistence and commercial livestock grazing systems are also at risk. In the protected areas, eco-tourism, which is major source of revenue for conservative efforts, also suffer as it creates an environment in which there is poor visibility for game viewing (Wigley et al, 2009). Other negative impacts of bush encroachment are the reduction in grassland diversity, at the ecosystem and biome scales, and subsequent loss of biodiversity, the lack of forage resources, the impeded movement and access of livestock to more palatable grass and provision of habitats for predatory wildlife (Dalle et al, 2006). The significant reduction or loss of the grass layer also alters regional fire regimes by reducing fire frequency and possible exclusion.

Bush encroachment; on the other hand, does not only impose harmful impacts on the environment and the local populace but can be considered beneficial depending on the land use practices and circumstances. According to Wigley et al (2009), bush encroachment can benefit the local communities by increasing the availability of wood for building, fencing and energy purposes (firewood). Certain bush encroaching species, such as *Terminalia sericea*, also serve as an important source of fuelwood for the local populace of the Bushbuckridge region near Kruger National Park (B. Erasmus, personal communication - April 2009). In a conserved or protected environment, bush encroachment can increase the food sources for the browsing animal populations and increases the diversity of bird species which utilise this woody vegetation for the construction of nests (Wigley et al, 2009). In some cases, soil erosion can be inhibited, especially in riparian zones and watersheds, by the presence of this woody encroachment whose roots anchor the soil and prevent excessive soil loss from surface run-off. Despite its dual perception of being both a blessing and a curse, bush encroachment is generally very difficult to manage and control as the

drivers or causes are not well known in scientific and management fields. It is, however, driven by a multitude of factors ranging from climate (amount of precipitation and seasonality), topography, soil texture, soil type and depth, and land management or land use practices (specifically grazing, patch dynamics, fire and wood harvesting) (Scholes & Archer, 1997, cited in Martin & Asner, 2005). Regardless, bush encroachment is an established phenomenon which persists in savanna eco-regions around the world. Within the context of Southern African savannas, the evidence of such persistence is best illustrated in the history of the iconic Kruger National Park which was summarised by Joubert (2007).

1.1.2 HISTORY OF BUSH ENCROACHMENT IN THE KRUGER NATIONAL PARK REGION

Understanding the history of bush encroachment, the historical environmental conditions and the specific land management strategies implemented in the Kruger National Park (KNP) region is paramount in understanding the current prevailing trends and patterns of this phenomenon. During the formative years of the KNP region, according to Stevenson Hamilton (1903) cited in Joubert (2007), the vegetation in the area south of the Sabi River was covered with very dense thorn bush with a lack of open country except in the extreme west. Bush, however, in most areas was sufficiently open to allow passage of horseman. From these observations, bush encroachment was virtually non-existent due to the annual veld burning practices (Stevenson Hamilton, 1912 cited in Joubert, 2007). During the 1926 to 1946 period, there were observations that there was a transition from primary forests (large trees) to secondary forests which contained stunted but fire-proof growth with an increased presence of woodiness. Also, the results of a rudimentary assessment pointed to the presence of bush encroachment in the Pretoriuskop and Tshokwane (southern and northern reaches of the park) (Joubert, 2007). During the 1946 to 1960 era, according to the KNP biologist's Annual Report of 1951, the worst presence of bush encroachment was in the Pretoriuskop region which was fronted by *Dichrostachys cinerea* and *Terminalia sericea* tree species. The increase in woodiness, mostly in the southern regions of the park, lead to a decrease in grazer populations while increasing the browser populations.

According to the Annual Report of 1963 and 1964, bush and shrub encroachment had increased at an alarming rate and had got to such a stage that the young trees and shrubs were too overwhelming in extent for elephants and other natural causes to regulate and remove them. Thus the only option to restore the equilibrium between trees, shrub and grass presence in the park was mechanical removal and chemical control which was met with limited success. A number of bush clearing campaigns were conducted in the southern bushwillow communities and in the northern mopane communities during the 1970's. Bush encroachment was cut back during the 1990, 1991 and 1992 seasons in which the severe drought conditions caused significant deaths of large encroaching tree species (Joubert, 2007). Although the combating of bush encroachment continued up till the present, it was recognised that complete and permanent eradication of the phenomenon was impossible as most of the encroacher seeds originated from sources outside of the park which were transported into the park via river systems (Joubert, 2007). In order to adequately manage this phenomenon, regional level monitoring must be considered as the way forward which can only be effectively achieved by remote sensing.

1.2 REMOTE SENSING AS A TOOL FOR MONITORING BUSH ENCROACHING SPECIES

The monitoring, assessment and analysis of environmental change, phenomena and trends at a regional geographical scale have always been the strength of the science and technology of remote sensing. This is because large extents of land, which would otherwise be logistically and physically impossible to survey at the ground level, could be analysed and interpreted from afar. Remote sensing would thus serve as the prominent tool for this study.

1.2.1 BUSH ENCROACHING SPECIES MAPPING USING REMOTE SENSING – A SHIFT IN PERSPECTIVE

Numerous studies, with special focus on the ones performed in the Kruger National Park and similar savanna environments, have been attempted to quantitatively study this associated woodiness of the bush encroachment phenomenon as a whole. These

quantitative studies made use of remote sensing techniques ranging from the fundamental supervised classification (Eckhardt et al, In. press) to textural analysis (Hudak & Wessman, 2001) to spatial connectivity of shrub patches (McGlynn & Okin, 2006) and to the more recent technique of Spectral Mixture Analysis (SMA) (Martin & Asner, 2005). On a personal note, my honours thesis study, similarly, focused on the mapping of this encroaching woodiness, using pixel-based supervised and unsupervised classification methods, across a 40 year time series of aerial photographs for the Pretoriuskop region in Southern Kruger Park (Naidoo, 2007 – BSc Honours thesis).

These studies help provide the different forms of land management (communal and protected) with the means to quantify the occurrence and presence of the bush encroachment phenomenon. Quantify the phenomenon itself can be very useful for general mitigation measures, but understanding the role of the different individual savanna tree species, which are known to be bush encroaching, in the various local land use practices is a challenge and is worth investigating. Such is the case in communal rangelands where certain woody savanna tree species are valued while others are considered a problem by the local communities depending on the prevailing land use activities. For example, *Terminalia sericea* is a valued fuelwood source for the local populace while other species such as *Dichrostachys cinerea* are known to be a nuisance to informal cattle herders as vital grazing pastures are lost. Thus understanding the distribution of these encroacher species will provide more useful information for managerial authorities than the traditional, simple quantification of the bush encroachment phenomenon. This shift in perspective would introduce a new wave of more species oriented land management strategies. Also, not all woody savanna tree species (e.g. *Euclea natalensis*) are necessarily bush encroaching species or contribute to bush encroachment and should not be subjected to the various mitigation measures put forward by management. Various challenges, however, are associated with mapping vegetation species at the individual species level, especially in a heterogeneous savanna environment.

1.2.2 THE CHALLENGES TO BE ENCOUNTERED IN THE MAPPING OF SAVANNA TREE SPECIES

Regardless of the type of environment, according to Hestir et al (2008), Lees & Ritman (1991) and Tong et al (2004), the ability and success of species mapping using remote sensing is directly and indirectly influenced by numerous factors. These include meteorological factors such as the prevailing weather, seasonality and sun viewing at the time of image acquisition. Physical factors such as the geology, edaphic conditions, topography and geography can also play a role in influencing species mapping. Biological factors, such as the existence of different plant phenological states that are dictated by the prevailing season, play a role in altering the intra-species spectral variability. The seedling, flowering and senescent phases of a plant or tree of the same species can possess different spectral response patterns which can make mapping difficult (Hestir et al, 2008 & Tong et al, 2004). Finally, the success of species level mapping using remote sensing is strongly dependent on the limitations of the sensor's spectral and spatial resolutions in relation to the size of the tree species canopies being mapped (Sobhan, 2007 – b & Tong et al, 2004). To accurately detect and map plant species, a sensor must have a very wide spectral range (high spectral resolution) and a high spatial resolution to collect sufficient spectral information and to sufficiently cover the particular tree species canopies for identification and mapping.

Numerous studies have readily dealt with the tackling and detection of different plant functional groups, like the mapping of broadleaf and fine-leaf forest trees (Kooistra, In. press) and different mangrove types (Yingchin et al, 2006), but very few studies have exclusively undertaken the identification and mapping at the species canopy level. Hestir et al (2008) investigated the identification of invasive aquatic species in the California Delta while Asner et al (2008 - b) tackled the discrimination and mapping of native and invasive tree species in the Hawaiian Rainforests and Sobhan (2007 – b) mapped tree and shrub species richness across a Mediterranean landscape but few have attempted this feat in an African savanna environment. This is due to the fact that these savannas are complex, heterogeneous environments with high intra-species variability due to the differences in geology (granite and gabbro), rainfall, herbivory and human impacts within relatively short distances (Cho et al, 2009 & Cho

et al, 2010). The savanna ecosystems are also highly irregular in shape and structural dimensions with a combination of open grassland patches and dense woody thicket – a stark contrast to the more homogeneous rainforest ecosystems. Vegetation structure and biochemical constituents (and thus leaf phenology) of these tree species are thus subsequently influenced by these environmental characteristics which alters their spectral response patterns. In order for the mapping of savanna tree species, especially bush encroaching species, to be considered possible, certain technological and methodological advancements in the remote sensing field are warranted.

1.2.3 RECENT DEVELOPMENTS IN REMOTE SENSING

The persistence of bush encroachment together with other land degradation problems over the decades have recently warranted a new era of technological advancements in the remote sensing field which would serve to critically dissect these phenomena for the better monitoring and management of natural environmental systems. According to Chambers et al (2007), such breakthroughs in remote sensing has transformed certain ecological disciplines as new insights on environmental gradients, species assemblages and ecologically relevant measures (such as moisture and nutrient content of canopies) have been generated. Sensors have became more sophisticated but versatile with the increase in spatial resolution, spectral resolution, number of spectral bands and spectral range. This meant that the imagery obtained covered a given area at a greater level of detail with an exponential increase in the spectral information available. This was possible as the electromagnetic spectrum was spectrally sampled at a greater frequency. These attributes were best typified with the transition from multi-spectral sensors to the advent of hyperspectral imagery. Hyperspectral imagery are acquired in hundreds of very narrow, contiguous spectral bands that can cover the visible, near-infrared, mid-infrared and shortwave infrared portions of the electromagnetic spectrum (0.4-2.5μm) (Lilliesand and Kieffer, 1994). This kind of imagery exceeds the capabilities of multi-spectral imagery in that hundreds of more channels of data are collected which allows for the construction of a nearly continuous reflectance spectrum for every pixel in the image. Within each pixel, the different spectral response patterns of the different ground cover components are represented (Lilliesand and Kieffer, 1994). Hyperspectral imagery enables the discrimination among and possible identification of the different ground

cover components or earth surface features present in the image which have absorption and reflectance characteristics that distinctly occur over these narrow bands (Lillesand and Kieffer, 1994). This level of spectral discrimination would not be possible with the coarse bandwidths available in multi-spectral imagery. With the emergence of hyperspectral imagery and the growth in the popularity of ecosystem structure related sensors (e.g. *Light Detection and Ranging* and *Radio Detection and Ranging*), a clearer understanding of ecosystem functions, in terms of biochemistry and structure, could be unlocked.

1.2.3.1 BIOCHEMICAL REMOTE SENSING OF VEGETATION

Hyperspectral imagery provides a better understanding of the particular spectral response patterns, spectral regions and possible predictable traits in the spectral reflectance and absorption curves of the numerous biochemical and physiological variables present in the environment. The level of variability of the biophysical and biochemical attributes of the particular tree species contribute to the spectral characteristics which can be used to distinguish between the different tree species in an environment (Gamon et al, 2005 cited in Asner, 2008 –c). Such biochemical constituents include plant pigments, water content, nitrogen and organic carbon. A study by Asner (2008 – c) made use of hyperspectral imagery to tackle rainforest canopy chemistry, physiology and biodiversity in order to identify the various sources of variability in spectral signatures. It was discovered that the spectral variability (and thus species mapping potential) at a leaf level was controlled by a multitude of biochemical and structural properties inherent in the foliage (Asner, 2008 –c & Clark et al, 2005). At the canopy level, spectral variability was generally controlled by tree structural features such as canopy structure and leaf volume etc (Clark et al, 2005). Another study by Asner et al (2008 – a) focused on tackling the spectral discrimination of native and invasive tree species, based on spectral characteristics, in the Hawaiian tropical rainforest. It was evident that the reflectance and absorption properties (signatures) of native trees were generally unique from the introduced alien trees (Asner et al, 2008 –a). Nitrogen-fixing trees also showed unique spectral signatures while there were subtle but significant spectral differences between highly invasive and passive introduced trees (Asner et al, 2008 –a). From these studies, it was clear that spectral differences related to the biochemistry of particular tree species

played a role in tree species discrimination but there was also an underlying structural component, especially at the canopy level (Asner, 2008 – c & Clark et al, 2005), which influenced these spectral characteristics and differences.

1.2.3.2 STRUCTURAL REMOTE SENSING OF VEGETATION

Since tree canopy biochemical properties in the hyperspectral reflectance signatures are greatly impacted by inter- and intra canopy structure, gaps and shadowing, it would not be wise to underestimate the importance of ecosystem structure in tree species mapping studies (Asner et al, 2007). Active sensors such as Light Detection and Ranging (LiDAR) can measure the structural components in the environment. LiDAR directly measures the three dimensional distribution of plant canopies and below canopy topography which gives very accurate estimations of vegetation height and canopy structure (Lefsky et al, 2002). The LiDAR technology will be elaborated upon in the sub-sections to follow. Instead of only supplying ancillary structural data, LiDAR data can be solely utilised in the mapping of tree species. A doctoral study by Kim (2007) made use of LiDAR intensity data and crown structure metrics to differentiate broadleaved and conifer tree species in the Washington Park Arboretum. The metrics calculated from the LiDAR point data, which detected individual tree species, indicated the tree size and the crown shape characteristics. According to Kim (2007), the intensity data is similar, in principle, to the spectral reflectance patterns of the particular species and may be considered useful in discriminating tree species. The results proved successful and LiDAR data could be used to differentiate the tree species at an individual species level and to discriminate between the conifers and broadleaved types with reasonable accuracies.

Because spectral and structural properties within an environment are inextricably linked, a new breakthrough sensor technology which combines both the spectral and structural sensing capabilities on a single platform was recently introduced.

1.3 CARNEGIE AIRBORNE OBSERVATORY (CAO) INTEGRATED SENSOR

The Carnegie Airborne Observatory (CAO) sensor is a unique sensor type with the integration of a hyperspectral sensor and a Light detection and ranging (LiDAR) sensor on a single platform. This airborne sensor platform was chosen over a space-borne platform as it captures the spatial heterogeneity of ecosystem change that could not be ascertained at such detail with ground-based or space-borne derived means (Asner et al, 2007). Also, according to the ERDAS field manual (ERDAS Inc, 1999), airborne sensor platforms are more flexible as it allows for more control over when and where the imagery data is gathered.

According to Asner et al. (2007), this CAO technology fills a niche in the remote sensing field and advances regional ecosystems research as ecosystem structure and chemistry could be directly probed and analysed. The hyperspectral sensor, which is a visible and near-infrared image spectrometer, can provide spectral information on environmental cover, the abundance of biological materials and the estimated concentration of biochemical constituents (nitrogen, water content and plant pigments etc.) while the waveform LiDAR can provide detailed structural information on environmental cover, ground topography, vegetation height and canopy shape, stratification and architecture (Asner et al, 2007). Since structural and spectral plant properties go hand-in-hand, this type of sensor integration will ensure greater success in the analyses of future ecological remote sensing research questions. For more information on the CAO system configuration and specifications, please refer to the Appendices 1A and 1B. This CAO integrated sensor was successfully implemented in the detection and mapping of different invasive species in the Hawaiian tropical rainforests (Asner et al, 2008 – b). This technology was also used to quantify fire fuels and the fire spread potential of invasive grasses in the Hawaiian Volcanic National Park (HVNP) (Varga and Asner, 2008). The CAO sensor, however, was never implemented in a heterogeneous savanna environment for savanna tree species mapping; until now.

Mapping perspectives and techniques that were limited by the spatial and spectral resolutions of available sensors can now be brought to the fore. From a personal

perspective, my journey of bush encroachment study has finally reached a suitable platform where high detailed sensory information and associated techniques are finally available to accurately and sufficiently tackle the phenomenon at the individual species level. It is thus important to note at this stage of the thesis that the focus is not on the mapping of the bush encroachment phenomenon itself (which would require multiple imagery and a time series analysis) but the mapping of bush encroaching savanna tree species.

1.4 OVERALL STUDY AIMS, OBJECTIVES AND THESIS STRUCTURE

1.4.1 AIM

The aim of this project is to spectrally decompose and describe the local patterns of savanna bush encroaching species across the Greater Kruger Park Region using the new CAO integrated system (Hyperspectral and LiDAR integrated sensor). The overall goal is to classify 23 of the most common savanna tree species (see Appendix 3A for list of species and their attributes) across the study region and then isolate the 6 main bush encroaching species (*Acacia gerrardii*, *Combretum apiculatum*, *Dalbergia melanoxylon*, *Dichrostachys cinerea*, *Pterocarpus rotundifolius* and *Terminalia sericea*) for study and interpretation. The Greater Kruger Park Region includes two different study regions with different land use practices and pressures: Sabi Sands region (conserved and protected to concession, privately owned land) and the communal rangelands of neighbouring tribal communities in the Bushbuckridge Municipal District (highly grazed and harvested).

1.4.2 OBJECTIVE

The overall objective of this project is to investigate the spectral and structural characteristics of the patterning of savanna tree species, particularly bush encroaching species, over regionally varying:

- Topography (crests, slopes and valleys) and drainage
- A more or less constant geology (gabbro substrates with small patches of granite)

- And different land use practices (subsistence farming and cattle ranching practices found in communal rangelands versus the protective management strategies adopted in national parks and private reserves)

This would lead to an understanding of the degree or severity of resource usage, fuelwood and grazing reserves and the need for land rehabilitation measures (mechanical or chemical removal of encroachers) in the entire region under investigation.

1.4.3 THESIS STRUCTURE AND OUTLINE

This thesis comprises of two main analytical chapters and a final synthesis chapter to conclude. Brief sections on the general study area (land use study regions) and general methodology (CAO data calibration and pre-processing) will follow this introductory chapter prior to the beginning of chapter 2.

1.4.3.1 CHAPTER 2

Chapter 2 is the first main analytical chapter which is entitled: '**The spectral discrimination of savanna tree species with emphasis on bush encroaching species using canopy level CAO spectra**'. This chapter covers the issue of exploring the spectral discriminatory potential between the different savanna tree species and if bush encroaching species are spectrally distinguishable from other non-encroaching tree species. This is achieved by the usage of various deterministic and stochastic spectral similarity measures.

1.4.3.2 CHAPTER 3

The next and final analytical chapter, Chapter 3, is entitled: '**The mapping of savanna bush encroaching species using multiple endmember Spectral Angle Mapper in a novel spectral and structural Decision Tree approach**'. This chapter addresses the actual mapping question and proposes a new Decision Tree approach to map savanna bush encroaching tree species. The decision tree architecture is tested against a traditional multiple endmember SAM classification approach and the design,

which yields the highest overall classification accuracy, is chosen for implementation and map creation.

1.4.3.3 CHAPTER 4

The final thesis chapter, Chapter 4, serves as a general synthesis chapter which draws in the overall conclusions of the entire thesis study based on the findings from Chapters 2 and 3. Various overarching shortcomings and recommendations, future scientific studies and ways forward are also discussed. Formal acknowledgements, dedications and appendices are included to end the thesis.

1.5 THE GENERAL STUDY AREA - THE GREATER KRUGER NATIONAL PARK STUDY REGION (L1-8)

The area under study in this project is located in the southern region of the Greater Kruger National Park in Mpumalanga and consists of three broad study regions. These study regions are Skukuza, located in the southern Kruger National Park, Sabi Sands Wildetuin, located adjacent to the western boundary of the Kruger National Park, and the Bushbuckridge Municipality District (BMD) which contain the informal community of the surrounding area. Skukuza is classified as conservation or protected land with Sabi Sands being classified as a combination of concession and privately owned conserved land. The BMD is generally classified as communal rangelands that are utilised by the livestock ranching of neighbouring tribal communities. The land use study site, L3 (not included in the study region map in figure 1.1), was positioned over the Marla Marla Private Game reserved area which was strictly restricted and access for field work was prohibited and was thus excluded from the thesis study.

1.5.1 BUSHBUCKRIDGE MUNICIPALITY DISTRICT (BMD)

The entire Bushbuckridge region is approximately 2600km² in area and extends into the southernmost portion of the Limpopo Province. Bushbuckridge was formerly part of the Bantustan districts of Mhala and Mapulaneng in Gazankulu and Lebowa respectively during the apartheid regime. Bushbuckridge later became a single rural

municipality after the democratic elections of 1994 (Giannecchini et al, 2007). In the west, near the Drakensberg escarpment, the mean annual rainfall is approximately 1200mm and decreases to 550mm in the east (in the flatter interior) across an approximately 100km transect (Shackleton, 2000). Because of the arid conditions and highly variable rainfall in the eastern parts, drought is a common occurrence (Shackleton, 2000). The general topography ranges from flat to undulating. The region supports two broad savanna vegetation types: Lowveld Sour bushveld (in the wetter western region) and Lowveld (in the drier east). With regard to the population dynamics of the area, human population density is generally high due to the settlement and housing policies from the apartheid government and the area has been receiving a significant influx of Mozambique refugees in the early 1980s (Giannecchini et al, 2007). In terms of land use zones, the main land use is commercial forestry in the west, high density rural/informal residential in the central parts and communal grazing in most parts of the region.

1.5.2 SABI SANDS WILDETUIN

The Sabi Sands Wildetuin is a private game reserve of approximately 54 000 hectares which is situated at 24º50'S and 31º30'E towards the western boarder of the central Kruger National Park (Ben-Shahar, 1991). The area possesses a summer rainfall regime with a mean annual precipitation determined by the 700mm isohyet in the south and 550mm isohyet in the north (Ben-Shahar, 1991). The terrain is similar to the BMD in that it is gently undulating with catena geomorphological sequences of crests, slopes and valleys. The dominating granite group which persists throughout the area belongs to the Nelspruit Granite Suite and are cut by dolerite intrusions in the form of dykes and large sills in the south western portions of the game reserve. These dolerite intrusions belong to the Timbavati Gabbro group with contrasting vegetation types. This soil type dominates the area covered in the L456 study region. The vegetation in the Sabi Sands Wildetuin is classified as general lowveld with dominant features such as woody vegetation with low canopies and dense tree stands with long straight trunks that occur in combination with open grasslands (Ben-Shahar, 1991).

1.5.3 SKUKUZA REGION, KRUGER NATIONAL PARK

The Skukuza region situated in the southern Kruger National Park, to the north of the Skukuza rest camp, was considered for study in this thesis. The Skukuza climate region supports the general Lowveld bushveld vegetation type and receives a mean annual summer precipitation of approximately 500-700mm (Landmann, In. press). The topography is also undulating with gentle slopes of 2º and 5º in angle (Shea et al, 1996 cited in Landmann, In. press). The underlying geology is dominated by large granite formations with few small but significant patches of gabbro. The granitic soils are sandy and frequently shallow and support heavily wooded savannas that are dominated by *Combretum* (Bushwillow) species with *Terminalia sericea* persisting in deeper sandy soils (Joubert, 2007). The gabbro derived soils primarily support tall tree varieties of *Acacia nigrescens* and *Sclerocarya birrrea* (Joubert, 2007). Like in the Sabi Sands Wildetuin, the level of grazing in the Skukuza region is highly variable and fluctuates with grazer and browser populations which are ultimately regulated by drought (mainly), predation and park management strategies such as culling practices.

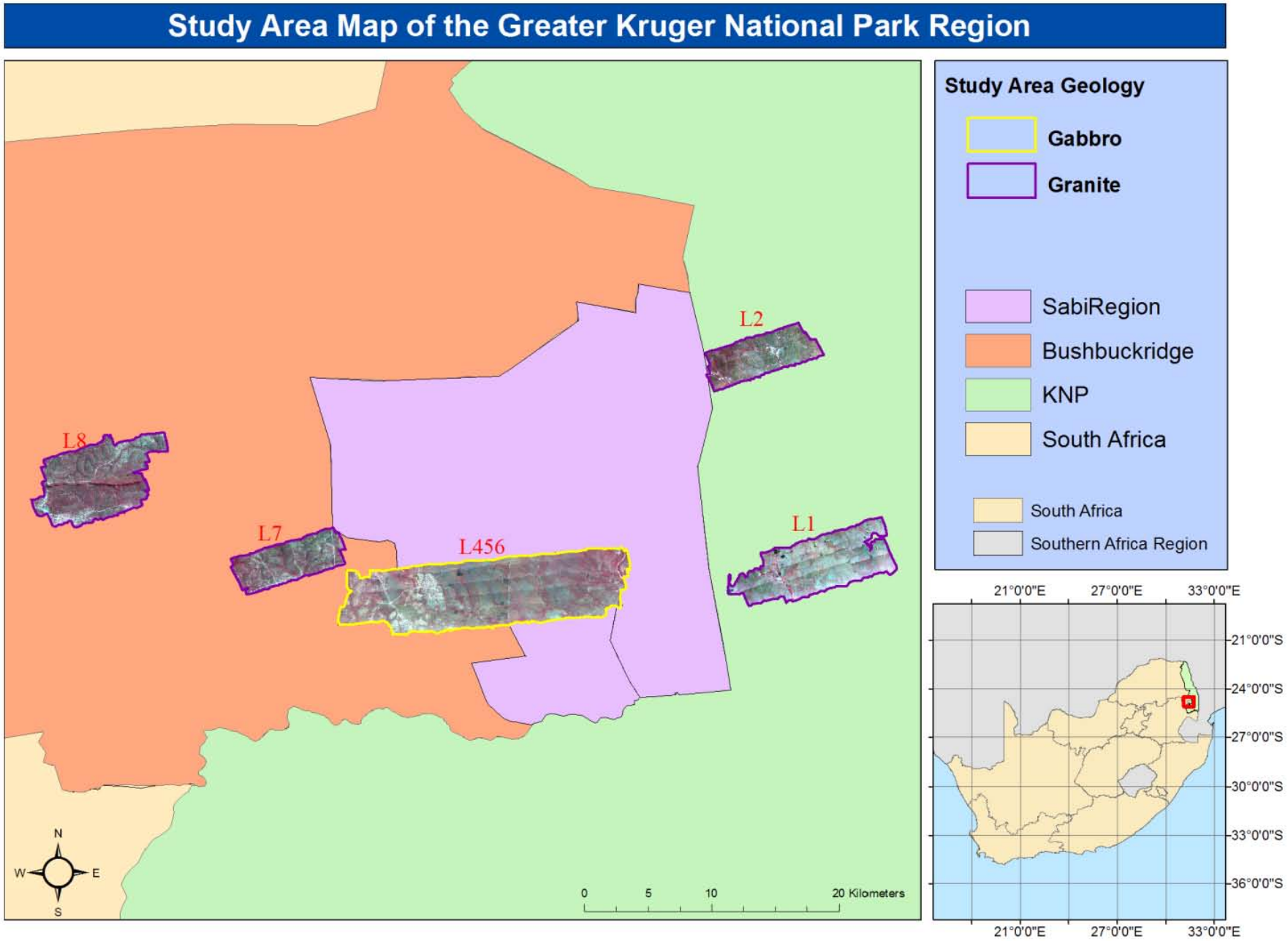


Figure 1.1: Map of the Greater Kruger National Park Study Area illustrating the different study regions (L1-L8) and their dominant geology

1.6 GENERAL THESIS MATERIALS AND METHODS

The subsections below indicate the general materials and methods that were used for the analyses conducted in the two main chapters (Chapters 2 and 3).

1.6.1 CAO DATA DESCRIPTION

At the end of the May 2008 CAO sensor flight campaign, the data products ascertained for this thesis study were the low resolution dataset (1m spatial resolution) consisting of five hyperspectral imagery and accompanying LiDAR vegetation height images of land use regions L1, L2, L456, L7 and L8 (displayed in the study area map in figure 1.1). These data products underwent various procedures of pre-processing and calibration before they could have been used in the two main analytical chapters. Accompanying this sensory data, various ground related data were collected during the April 2009 and May 2010 field campaigns which were discussed in the relative chapters in which they were utilised.

1.6.2 PRE-PROCESSING AND CALIBRATION OF CAO DATA

The hyperspectral images underwent a conversion procedure in which the raw digital number (DN) measures were converted to radiance and then to relative surface reflectance measures. The initial conversion to at-sensor radiances was achieved by applying radiometric corrections developed during sensor calibration in the laboratory. Apparent surface reflectance was then derived from the radiance data using an automated atmospheric correction model, ACORN 5LiBatch (Imspec LLC, Palmdale, CA). Inputs to the atmospheric correction algorithm included surface elevation (captured from the LiDAR), aircraft altitude (from GPS-IMU), solar and viewing geometry, and estimated visibility (in km). The code used a MODTRAN look-up table to correct for Rayleigh scattering and aerosols. Water vapour was then estimated directly from the 940 nm water vapour feature in the radiance data. This atmospheric correction and conversion from DN to surface reflectance unlocks the full potential of the images for spectral analyses procedures.

For the LiDAR data, the GPS-IMU data were combined with the laser ranging data to determine the 3-D location of the laser returns. From the laser "point cloud" data, a physical model was used to estimate surface and ground models (Digital Surface Model and Digital Ground Model). From this LiDAR data, a tree height map was created which was instrumental for the masking procedures in Chapters 2 and 3.

CHAPTER 2: SPECTRAL DISCRIMINATION OF SAVANNA TREE SPECIES, WITH EMPHASIS ON BUSH ENCROACHING SPECIES, USING CANOPY LEVEL SPECTRA FROM THE CAO HYPERPSECTRAL DATA

OVERVIEW

Before the mapping of savanna tree and bush encroaching species in the Greater Kruger Park Region is considered possible, one needs to understand the spectral differences between the bush encroaching and savanna tree species in order to spectrally discriminate the various species from one another. Spectral similarity measures have been utilised in remote sensing to specifically address the task of species spectral discrimination. Spectral similarity measures statistically compares the spectral properties between different tree species to ascertain how spectrally similar one species is to another. This chapter aims to utilise three spectral similarity measures – Spectral Angle Mapper (SAM), Spectral Information Divergence (SID) and the Spectral Information Divergence-Spectral Angle Mapper (tan) combination measure [SID-SAM (tan)] – to spectrally discriminate 49 savanna tree species using the CAO hyperspectral imagery. These similarity measures were also performed across 14 different spectral scenarios to determine the spectral region or band combinations most useful for species discrimination. The SAM measure was found to yield the best overall result for discriminating most of the species under investigation, especially spectral similar species and the various bush encroaching species. The bulk of the species discriminatory potential found in the SAM results were present in the transformed entire spectrum and primary and secondary plant chemical wavelength spectral scenarios. These results provide an insight into the classification algorithms and spectral regions needed for the effective mapping of savanna bush encroaching species which will be addressed in the next thesis chapter.

2.1 INTRODUCTION

2.1.1 THE IMPORTANCE OF TREE SPECIES DISCRIMINATION

In order to understand bush encroachment at the finest of scales one needs to understand the spectral properties and, ultimately, the spectral discriminatory ability of these associated bush encroaching species in relation to the surrounding savanna trees species. Once the species spectral separability has been established for the necessary savanna species, species level mapping and monitoring of their spatial distribution can be made possible. The most important motivation for plant species discrimination is for the proper sustainable management of any ecosystem because through the understanding of species composition and distribution, it will provide vital knowledge of the associated environmental changes (such as degradation) and the natural succession occurring within a particular ecosystem (Nagendra, 2000 cited in Sobhan, 2007 – a). The high spatial and spectral resolution of airborne hyperspectral sensors should facilitate the species discrimination process and would be more time-efficient than the older techniques of floristic mapping and extensive field surveys.

2.1.2 VEGETATION INDICES AND STATISTICAL MEASURES

Various statistical measures and indices have been utilised in remote sensing to provide quantitative information to address various issues such as carbon stock inventories, temporal land cover change and impacts of human activities (Lyon et al, 1998). In vegetation related studies, spectral vegetation indices have been commonly used to predict and quantify various ecological variables such as percentage vegetation cover, above ground biomass and leaf-area index (Lawrence and Ripple, 1998). One of the most common types of vegetation indices, based on brightness values, is ratio-based indices. According to Lawrence and Ripple (1998), ratio-based indices make use of the chlorophyll absorption characteristics of vegetation in the red portion and the high reflectance of vegetation in the near-infrared portions of the spectrum (e.g. Normalised Difference Vegetation Index – NDVI). Other types incorporate the soil component in their properties such as Soil-adjusted Vegetation Index (SAVI). These different types (ratio- and soil-based) of vegetation indices were evaluated in the Lyon et al (1998) study according to their performance, statistical characteristics and their ultimate capabilities in distinguishing land surfaces, water surfaces and cloud covers for change detection analysis. Other studies such as Pu et al (2008) made use of NDVI differencing and

thresholding to monitor sparse vegetation coverage. A study by Thenkabail et al (2000) evaluated the performance of various types of hyperspectral vegetation indices (NDVI type indices, soil adjusted indices and optimum multiple narrow band reflectance – OMNBR – models) and determined the spectral bands best suited for studying agricultural crops based on their biophysical variables. For broad levels of spectral discrimination (i.e. of land surface types), such indices and statistical measure may have proven adequate but for more specific levels of the spectral discrimination (i.e. of numerous, different vegetation or tree species), the use of spectral similarity measures was considered.

2.1.3 SPECTRAL SIMILARITY MEASURES

The most effective method to address tree species discrimination involves the implementation of spectral similarity measures. These spectral similarity measures statistically compares the spectral properties (as a result of surface composition) between different materials or targets (e.g. tree species) to ascertain how spectrally similar one material is to another (Van der Meer, 2006; Du et al., 2004). This vital information of material discriminating ability will facilitate material or target identification and subsequent mapping at the landscape scale. The use of spectral similarity measures is favourable in such a hyperspectral sensor application rather than the use of different vegetation indices, which also compares the specific spectral differences. This is because, as in the case of the vegetation indices, the data dimensionality and thus spectral information is reduced during processing (Sobhan, 2007 – a). Spectral similarity measures utilise the full available spectral band range and thus maintains full data dimensionality. According to Sobhan (2007 – a), the discriminating power of a hyperspectral sensor is ultimately limited by the spectral variability that occurs within species (intra-species) and between different species (inter-species). If inter-species spectral variability exceeds the intra-species spectral variability, then species discrimination is possible. The biochemical and leaf properties (and thus associated absorption and reflectance characteristics) are usually known to have a wider variation between different species than within species (Asner et al, 2008 – c; Sobhan, 2007 – a). According to Cho et al., (2009), intra-species spectral variability could result from differences in geological substrates on which the species occur, rainfall, herbivory and fire pressures plus differences in human activities (being harvested in a communal setting or protected in a national park setting).

2.1.3.1 SPECTRAL SIMILARITY MEASURE THEORY AND RELATED STUDIES

There are two categories of spectral similarity measures which exist in the literature – deterministic (empirical) measures and stochastic measures (Van der Meer, 2006; Du et al., 2004 & Sobhan, 2007 –a). Deterministic or empirical measures utilise the geometric properties of the spectra such as the spectral angle and the Euclidean distance of the spectral vectors (Sobhan, 2007 – a & Van der Meer, 2006). Deterministic measures include spectral similarity measures such as the Spectral Angle Mapper (SAM) and the Euclidean Distance (ED) (Du et al., 2004). Stochastic measures consider the spectrum of a pixel vector as a probability distribution and consider the spectral band-to-band variability in terms of uncertainty (Sobhan, 2007 – a & Du et al., 2004). The Spectral Information Divergence (SID) is an example of a stochastic measure (Van der Meer, 2006). The more commonly used deterministic and stochastic similarity measures; which include SAM, ED, SID and Spectral Correlation Measure (SCM); will be critiqued next according to their properties and applicability to this chapter study.

SID handles the spectrum of a pixel vector as a probability distribution and characterizes the spectral properties (such as the mean and variance) associated with a particular pixel vector (Du et al, 2004). SAM, on the other hand, measures the angle between two spectra, which are considered as vectors in n-dimensional space, to determine the degree of spectral similarity between those spectra (Cho et al, 2009; Sobhan, 2007; Du et al, 2004 & Van der Meer, 2006). The smaller the angle between the two spectra, the more spectrally similar the spectra are while the converse is true if the angle is larger. If the angle produced is very small, the SAM measure is known to represent the ED between the two spectra i.e. the distance between two spectra as a function of spectral similarity (Du et al, 2004). SAM proves to be advantageous as it is not affected by albedo effects, imagery cross-track effects and illumination associated with the different spectra (Kruse, 1997 cited in Sobhan, 2007 –a). ED takes into account the differences in brightness between the pixel vectors (Van der Meer, 2006). The ED similarity measure was not considered as an appropriate similarity measure as it is derived by and thus biased towards the SAM measure. In a complex, heterogeneous savanna environment, differences in brightness would not provide enough spectral information to discriminate between the different savanna tree species. The next measure, SCM, calculates the cross-pixel vector correlation by examining the correlation co-efficient of a pixel and the associated spectral signatures (Van der Meer, 2006). The SCM is sensitive

to the overall shape of the reflectance curve (spectrum) which can be significantly altered when the spectral bands are reduced to the significant bands for species discrimination (Sobhan, 2007 – a & Van der Meer, 2006). This is because it can no longer make use of the overall differences in the shapes of the curves. For this reason SCM was excluded from this study. The performance of these 4 spectral similarity measures (SID, SAM, ED & SCM) were evaluated, by Van der Meer (2006), in a geological scenario which focused on discriminating 4 dominant geological minerals. In order to determine which measure yielded the optimum results, a Probability of Spectral Discrimination (PSD) was utilized. This measure simply creates a comparative statistic to compare the different spectral similarity measures which have different value ranges. This measure will be discussed in greater detail in the validation section of this chapter's methodology. The overall results from the Van der Meer (2006) study showed that the SID measure was significantly the most effective for discriminating the different geological minerals while SAM was the least effective. It was also important to note from this study, that the different measures' separability values improved when the measures were focused on a few important wavelengths rather than the entire available spectral range.

A new spectral similarity measure known as the SID-SAM mixed measure was introduced by Du et al (2004) to challenge the performances of the traditional and more established measures, such as SAM and SID, in discriminating 5 different materials (shrubs, leaves and red soil). Another study, by Sobhan et al (2007 – a), made use of the SAM, SID, SCM and this SID-SAM mixed measure to explore the spectral dissimilarity between different main herbs, shrubs and tree species in a Mediterranean landscape. According to Du et al (2004), the SID-SAM mixed measure incorporates the advantageous properties of both the SID and SAM measures and makes similar spectral signatures even more similar while making dissimilar spectral signatures more distinct. This measure is known to perform well when the interspecies spectral variability is high. The outcomes of both the Du et al (2004) and Sobhan et al (2007 – a) studies illustrated that the SID-SAM mixed measure yielded the best discriminatory ability results according to the corresponding high PSD values. From the Sobhan et al (2007 – a) study, it was also clear that the spectral similarity was significantly higher within a given species than that between the different species with the discrimination being most evident between species types (e.g. herbs, shrubs and trees).

Other measures such as the Mahalanobis Distance was not utilised in this study as it was fairly intensive in that it required a large amount of sample training data (compared with data dimensionality, it is mostly not enough) and that it possessed high computational costs (Sun et al, 2000). Other studies, from Vaiphasa et al (2005) and Clark et al (2005), utilized a combination of purely statistical approaches to investigate the spectral discriminatory ability between mangrove and tropical tree species. These statistical approaches which utilised a combination of techniques such as a one-way ANOVA, Jeffries-Matusita distance measure and non-parametric multivariate ANOVA, were not considered in this thesis chapter as spectral similarity measures, plus their accompanying validation measures, were considered simpler to compute. This was because there was no need for combining other various statistical measures to achieve a viable result. Despite the various techniques and studies which have dealt with the concept of vegetation species discrimination, very few studies have actually attempted to apply these techniques to the spectral discrimination of savanna tree species at the canopy level. This study seeks to help fill this gap in the literature.

2.1.4 CHAPTER AIM AND RESEARCH QUESTIONS

The aim of this chapter is to spectrally discriminate bush encroaching species at a canopy level in comparison to other savanna species, using the CAO hyperspectral imagery data of the Greater Kruger National Park Region (Southern Kruger, Sabi Sands Wildetuin and Bushbuckridge municipality district). The spectral similarity measures that will be tested for species discrimination of savanna and associated bush encroaching tree species are SAM, SID & the SID-SAM mixed measures. These measures were selected for the analyses in this chapter as they performed effectively when it came to studies relating to plant species discrimination and separability. With this in mind, various related questions need to be answered:

- Which spectral region is or wavelength combinations are most suited for species discrimination?
- Does primary and secondary plant chemical spectral absorption wavelengths aid in this discrimination?
- Can the continuum removal transform function be used to enhance the absorption features necessary for species discrimination?

- Are there enough spectral differences to spectrally differentiate the bush encroaching species from the other non-encroaching savanna tree species?
- Which spectral similarity measure yields the optimum results?

Approximately 49 savanna species (including unidentified species) were analysed in this chapter with special focus on 8 bush encroaching species, *Combretum apiculatum, Dichrostachys cinerea, Pterocarpus rotundifolius, Strychnos madagascariensis, Strychnos mozambiquensis, Terminalia sericea, Acacia gerrardii* and *Acacia exuviales*, which were dominant encroaching species across the study regions (Joubert, 2007). The associated results, from the above questions, will prove invaluable for the mapping process that will be addressed in the next chapter of this thesis. The results will also be compared in the light of the results from other pertinent studies.

2.2 METHODOLOGY

This methodology section consists of sub-sections pertaining to the collection of field data collected from the April 2009 field campaign, the implementation of this data in three spectral similarity measures under various spectral scenarios and the necessary validation protocols.

2.2.1 APRIL 2009 FIELD CAMPAIGN

A joint CAO-CSIR field campaign took place during the month of April 2009 in which the Southern Kruger National Park (KNP), Sabi Sands Wildetuin and the Bushbuckridge municipality district study regions were sampled. A spectral library containing leaf level spectra, collected from the field, and pixel level spectra, collected from the CAO imagery, was created. This spectral library contained as many savanna tree species as possible that were encountered in the field.

Four transects were made across the study sites (hyperspectral image scenes) that covered the entire study region with 1 transect across the study site in KNP, 1 transect across Sabi Sands Wildetuin and two transects across the communal rangelands of the Bushbuckridge municipality district. Each transect contained a number of sample points that were spaced every ~300m intervals with the number of sample points varying according to the length of

the transect walked. Along the transect, very large canopy trees (greater than ~2m canopy diameters) were focused on as they would have been clearly visible in the CAO imagery, taken in the previous year. These adequately sized tree canopies would be ideal for extracting spectral endmembers for each species. At each large tree, the species and GPS positioning was recorded. Along the transects and at the individual sample sites, an ASD spectroradiometer was used to collect leaf level spectra for the tree species encountered and for different soil colour types. The leaf level spectra were measured from healthy leaves which were plucked off accessible tree branchlets. A total of 176 trees were recorded, with GPS, which comprised of approximately 49 different tree species. The field data collected along the transect was vital for the final construction of the spectral library. The leaf level spectra collected from the ASD spectroradiometer was not used for the analyses in this chapter. This was because pixel level spectra would be more suitable for endmember extraction as it would be representative of the entire tree canopy, rather at a localised leaf level of a few sample leaves, for savanna tree species classification.

2.2.2 SPECTRAL DATA PRE-PROCESSING

During the data processing phase, pure pixels (reference spectra) of the 49 savanna tree species (including the bush encroaching species) were extracted at canopy level over the Visible and Near-Infrared spectral regions of the imagery. The CAO low spatial resolution (1m) imagery was used for this study as it covered a larger geographical area than the high resolution imagery and it allowed the overlapping of all the data collected from the field. This pure pixel extraction process was achieved by overlaying GPS derived point shapefiles of the trees (including species information) over the hyperspectral imagery and using the spectra profile tool (in ENVI 4.5 remote sensing software) to collect only the pure pixels of each species' canopies. This process was repeated for all GPS positioned trees, collected from the field data, across the hyperspectral imagery in which the field transects were conducted. The spectral reflectance values and the band wavelength information for the numerous pure pixel samples for each species were compiled into a spectral library in Microsoft EXCEL. The mean spectral values of the reference pixels were then calculated for each species. These mean reference spectra for each species were then applied to three different spectra similarity measure formulae on a pairwise basis – SAM, SID and SID-SAM mixed measure - to determine species separability. SAM and SID Matlab scripts were

created to run these spectral similarity formulae in a quick and efficient manner. The entire SAM and SID scripts are displayed in Appendices 2D and 2E.

2.2.3 SPECTRAL SIMILARITY MEASURE FORMULAE

Just to reiterate, SAM calculates the spectral similarity by measuring the angle between the spectral signature of two samples (in this case, species) (Yuhas, 1992; Kruse et al, 1993 cited in Sobhan, 2007; Du et al., 2004). The narrower the angle between the two species spectra, the more spectrally similar they are to one another (Sobhan, 2007). It is represented by the formula below:

$$SAM(sp1;sp2) = cos^{-1}\left(\frac{\sum_{l=1}^{L} sp1l \times sp2l}{[\sum_{l=1}^{L} sp1^2]^{\frac{1}{2}}[\sum_{l=1}^{L} sp2^2]^{\frac{1}{2}}}\right) \quad (1)$$

Where sp1 and sp2 refers to the particular reflectance values (at a particular band) for a particular tree species 1 and 2. L refers to the total number of available bands.

SID calculates the distance between the probability distributions produced by the spectral signatures of two pixels (Van der Meer, 2006). This is represented by the formula below:

$$SID\ (sp1;sp2) = D(sp1 \parallel sp2) + D(sp2 \parallel sp1) \quad (2)$$

$$where\ D(sp2 \parallel sp1) = \sum_{l=1}^{L} ql[Il(sp1) - Il(sp2)]$$

$$and\ D(sp1 \parallel sp2) = \sum_{l=1}^{L} pl[Il(sp2) - Il(sp1)]$$

Where p and q are probability vectors for the spectral signatures of species 1 and species 2. Il represents the natural log or –log. L refers to the total number of available bands.

The SID-SAM mixed measure encompasses the combination of the properties of both SAM and SID similarity measures and makes two similar spectra more similar and two dissimilar spectra more spectrally distinct (Du et al., 2004 & Sobhan, 2007). The SID-SAM mixed measure is expressed and applied either in a sine (sin) or tangent (tan) version (Du et al., 2004). The tangent version was used in this chapter's analyses as it produced marginally

higher similarity values than its sine counterpart. This SAM-SID mixed measure is represented by the formula below:

$$\text{SID-SAM(tan)} = \text{SID(si, sj)} \times \tan[\text{SAM(si, sj)}] \quad (3)$$

Where si,sj refers to the SID or SAM value for the pair-wise comparison between a particular species 1 and species 2 in the similarity measure matrix.

2.2.4 SPECTRAL SCENARIOS

Before the mean reference pixel values, for each species, were applied to the similarity measure formulae, the spectral range of the spectral library data were divided into the different spectral scenarios to see which region or wavelength combinations possessed the greatest species discriminating power. It is important to note that these different spectral scenarios were compiled from the closest available wavelengths present in the species spectral library. These spectral scenarios included the entire spectrum (394.3nm to 1044.9nm), the red region (602nm to 696.6nm), blue region (403.7nm to 498.1nm), green region (507.5nm to 592.6nm), the visible spectrum (403.7nm to 696.6nm), near-infrared region (800.5nm to 998.1nm), spectral wavelengths in which primary and secondary plant chemicals, such as chlorophylls a-b; carotenoids & xanthophyll, were most expressive (the following wavelengths: 422.6nm; 432nm; 441.4nm; 450.9nm; 469.7nm; 479.2nm; 488.6nm; 639.9nm; 658.8nm; 932.3nm; 988.7nm & 1044.9nm) and finally their continuum removed (transformed) counterparts. Please refer to the table in Appendix 2B for the different wavelengths and the various primary and secondary chemicals which they express. Subjecting the spectral scenarios to a form of transformation, continuum removed, was a very important step in the analyses. This transformation was done to enhance the absorption features of the mean reference spectral values evident in the spectral profiles and to minimize the differences caused by the variability of solar illumination at each pixel-crown position (Odagawa & Okada, In. press). This transformation thus ultimately attempts to create better species discrimination results. The output of these similarity measures is similarity matrices which display the degree of similarity between the different tree species in comparison to the other species within each spectral scenario. The total SID, SAM and SID-SAM values for each species and the overall SID, SAM and SID-SAM values for the entire matrix were calculated and compared. These results were then graphically represented in the form of column graphs for comparisons across species and across the different spectral scenarios.

2.2.5 VALIDATION

Since these different spectral similarity measures produced results of different scales and value ranges, this study made use of a validation measure which created a comparative statistic to ultimately assess the performance of the three different spectral similarity measures in order to find out which measure yielded the most plausible result. The validation measure known as the Probability of Spectral Discrimination (PSD) was used in this study. The PSD calculates the relative capability of all spectra to be discriminated from others (Sobhan, 2007). This means that the higher the probability, the better the capability of a species to be discriminated from others (Sobhan, 2007). In terms of species mapping, PSD assesses the ability of a set of reference spectra or endmembers to map a particular species (Van der Meer, 2006). The PSD formula, used in this study, is displayed below.

$$PSD \;\; Pt, \Delta(k) = \frac{m(t, sp1)}{\sum_{j=1}^{L} m(t, sp2)}, \quad for\, k = 1, \ldots . K \tag{4}$$

Where m is any of the pre-defined similarity measure and $\sum_{j=1}^{L} m(t, sp2)$ *is defined as the sum of all similarity measures in the matrix. m(t, sp1) is the sum of all similarity measures for the target spectrum t relative to other spectra (Van der Meer, 2006). K refers to the number of mean spectral endmembers (i.e. species) implemented from the spectral library.*

The SAM, SID and SID-SAM (tan) output matrix values were applied to the PSD formula and were analysed at a species level to see which measure yielded the best overall discriminating results.

2.3 RESULTS

In order to best represent the results from the SAM, SID and SID-SAM matrices, the total similarity measure values for all 49 species (refer to Appendix 2C for the scientific names and their associated abbreviated names) was compared only across the three major spectral scenarios which illustrated the most representative results. These results took the form of column graphs which were the most appropriate form of representation as the overall discriminatory ability of each species in relation to the rest of the species (from the total SAM, SID or SID-SAM value) could be compared. From these column graphs, the performance of the measures within all the spectral scenarios/regions for species spectral

discrimination could also be assessed (i.e. both at an individual species level and overall spectral scenario level). The PSD values for every spectral scenario were pooled together on a per species basis for each spectral similarity measure and the overall PSD species values were plotted on a column graph for comparison.

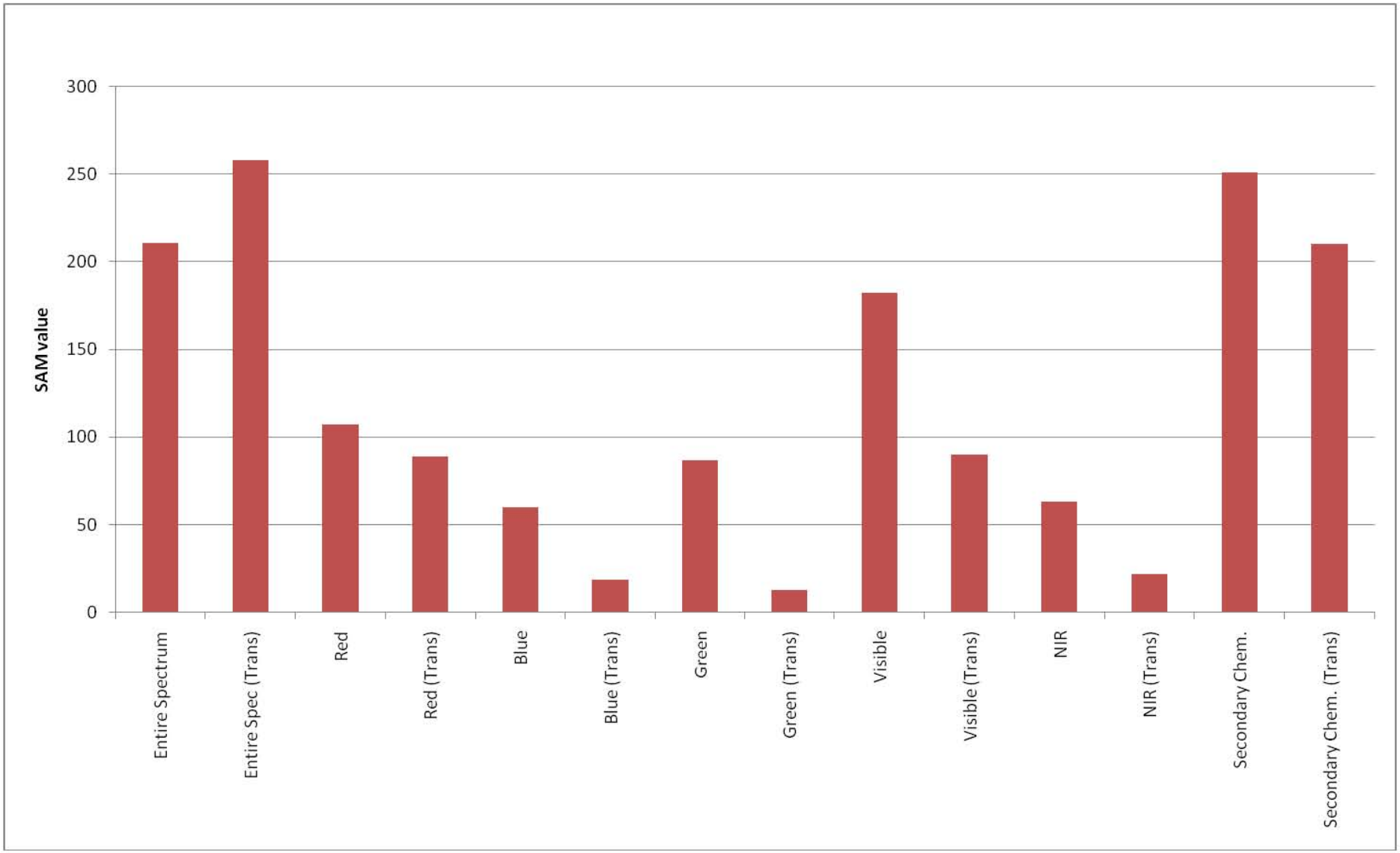


Figure 2.1: Overall accumulated SAM values across all 14 spectral scenarios

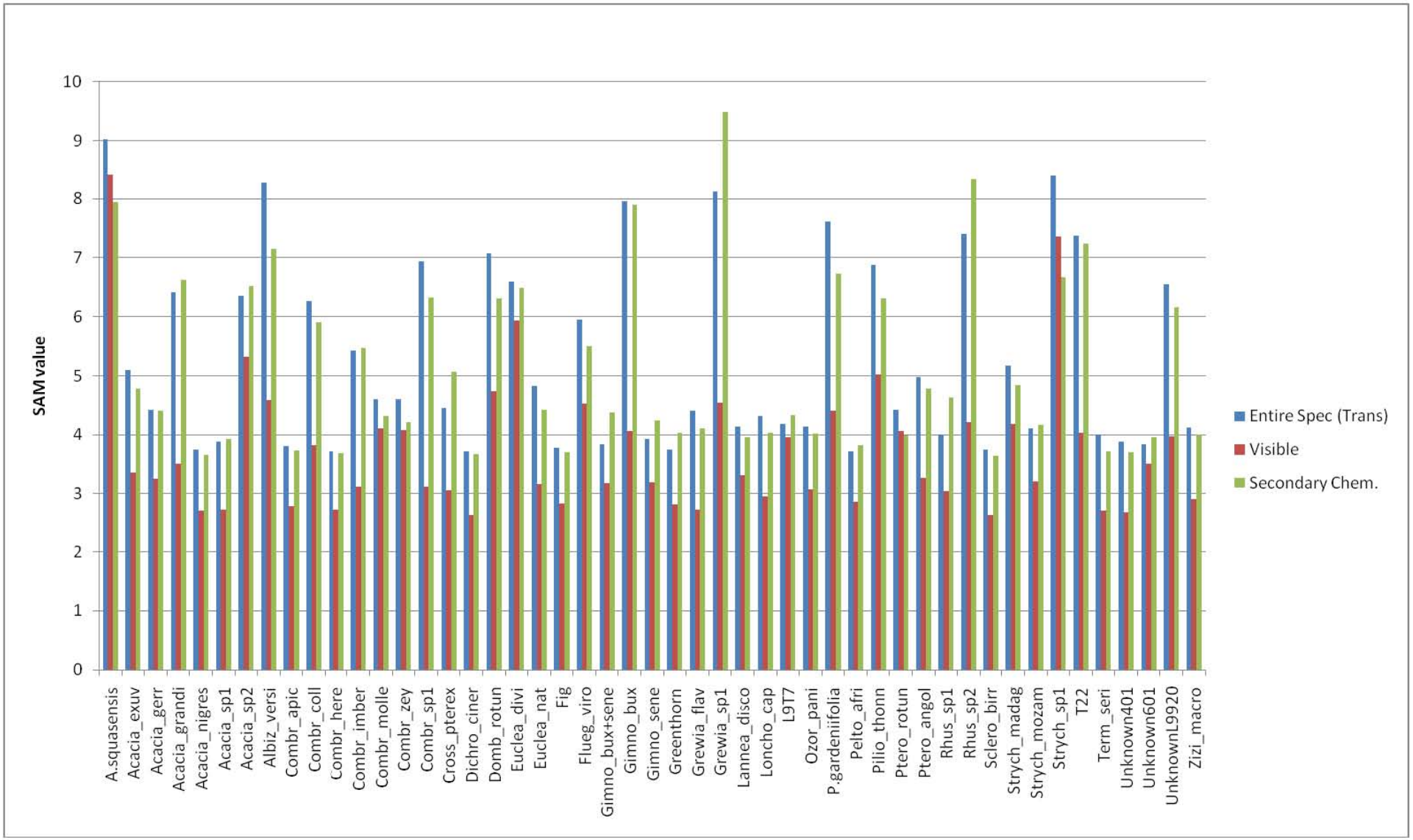


Figure 2.2: SAM values for different species across 3 dominant spectral scenarios

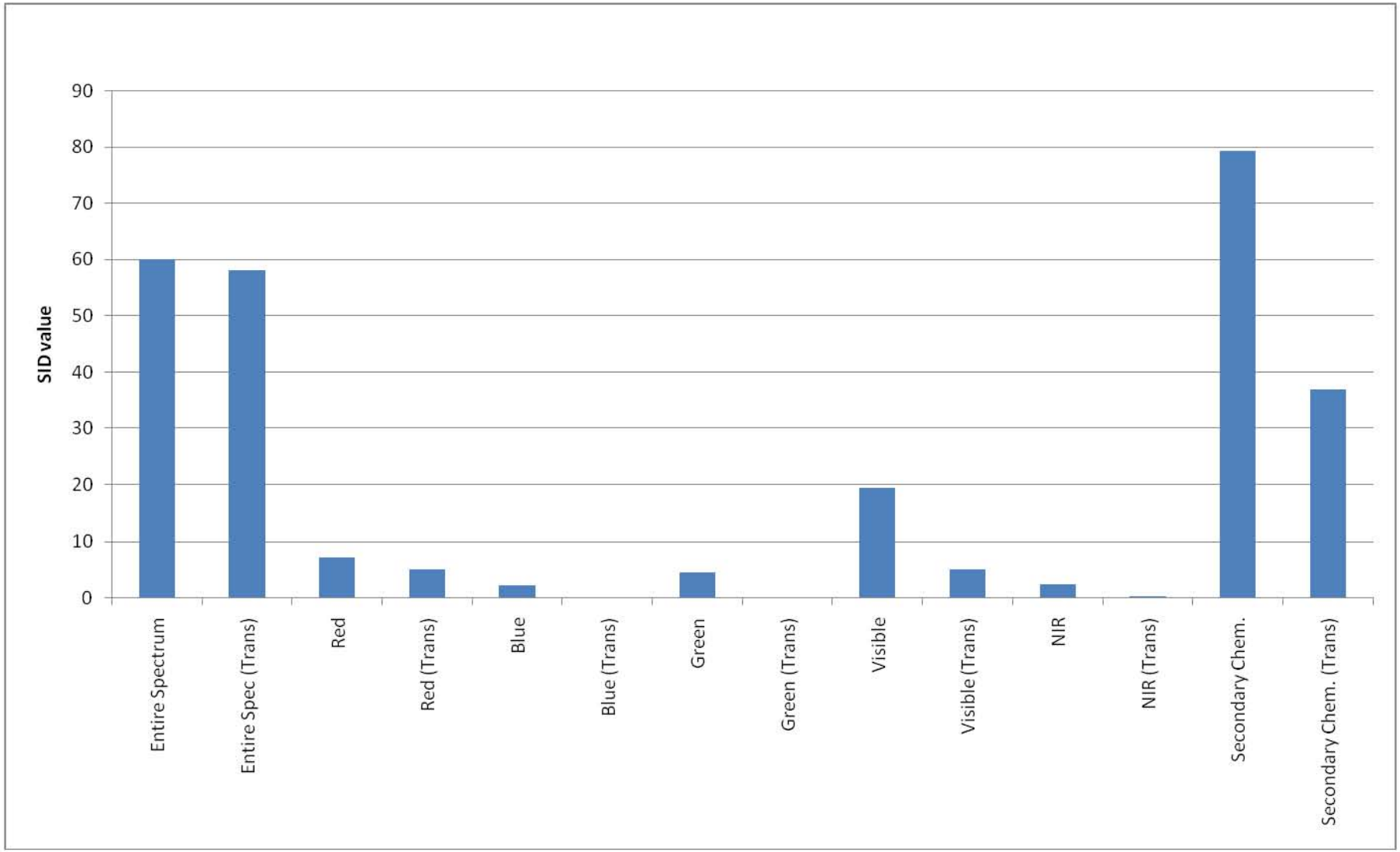


Figure 2.3: Overall accumulated SID values across all 14 spectral scenarios

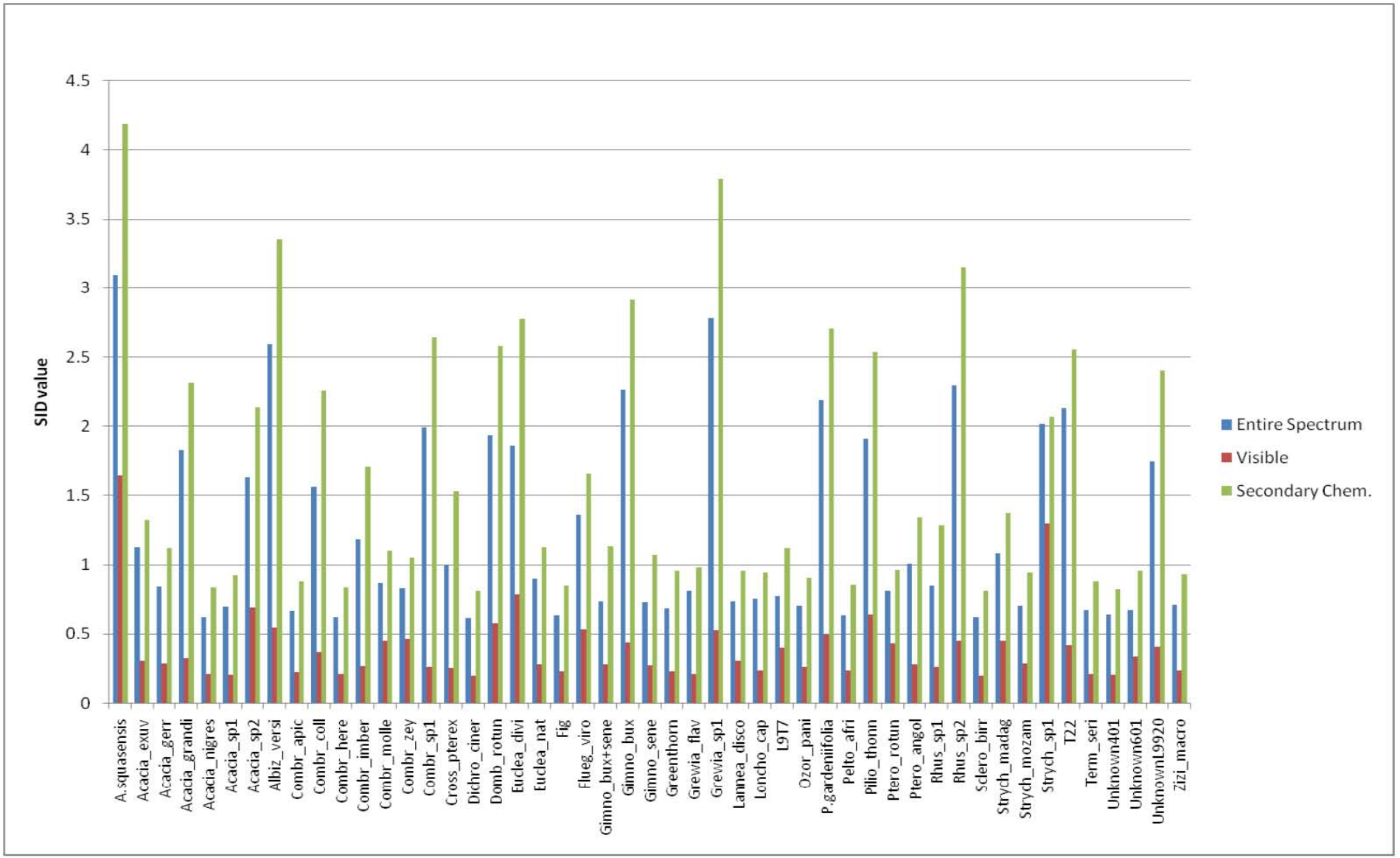


Figure 2.4: SID values for different species across 3 different spectral scenarios

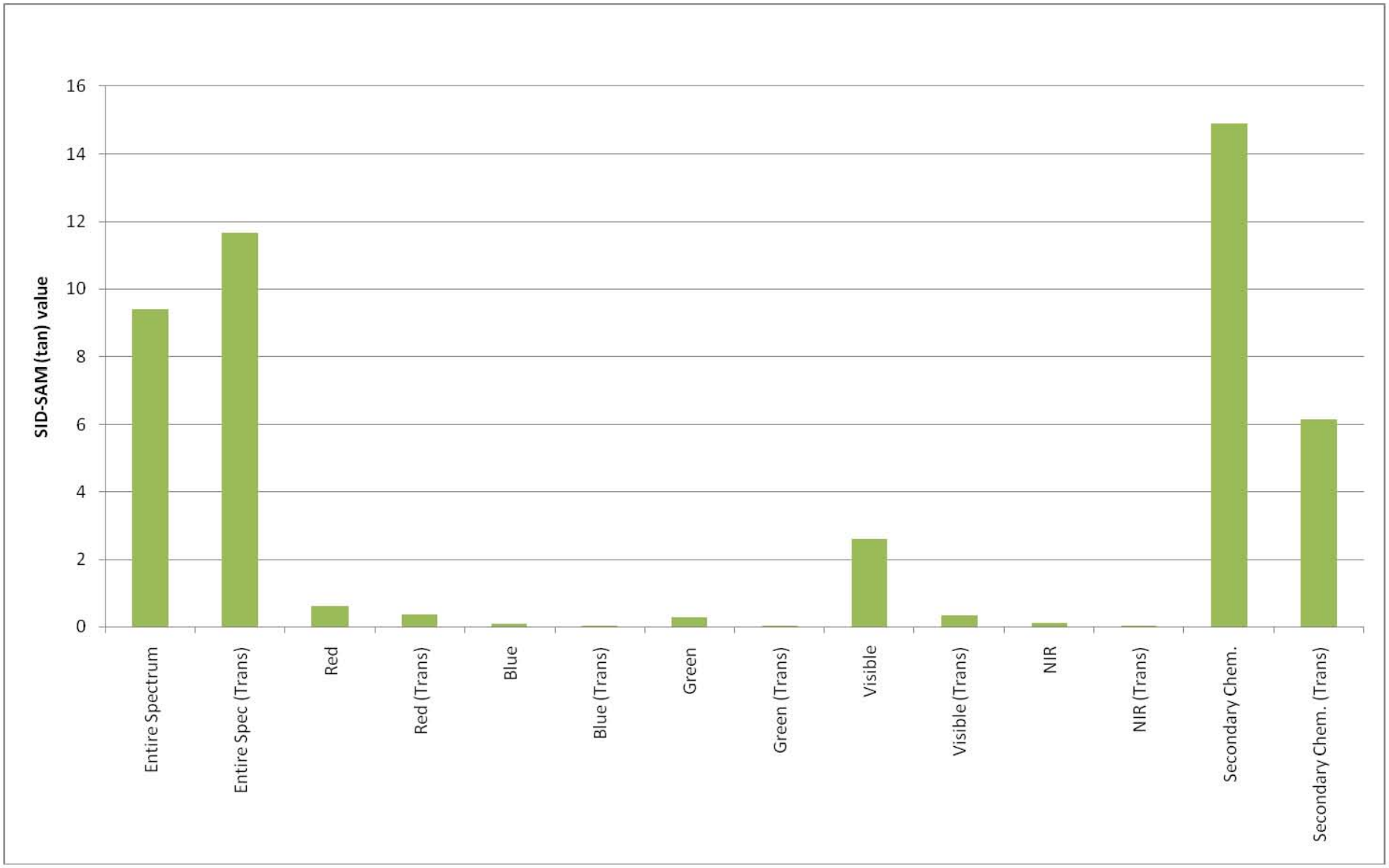


Figure 2.5: Overall accumulated SID-SAM (tan) values across all 14 spectral scenarios

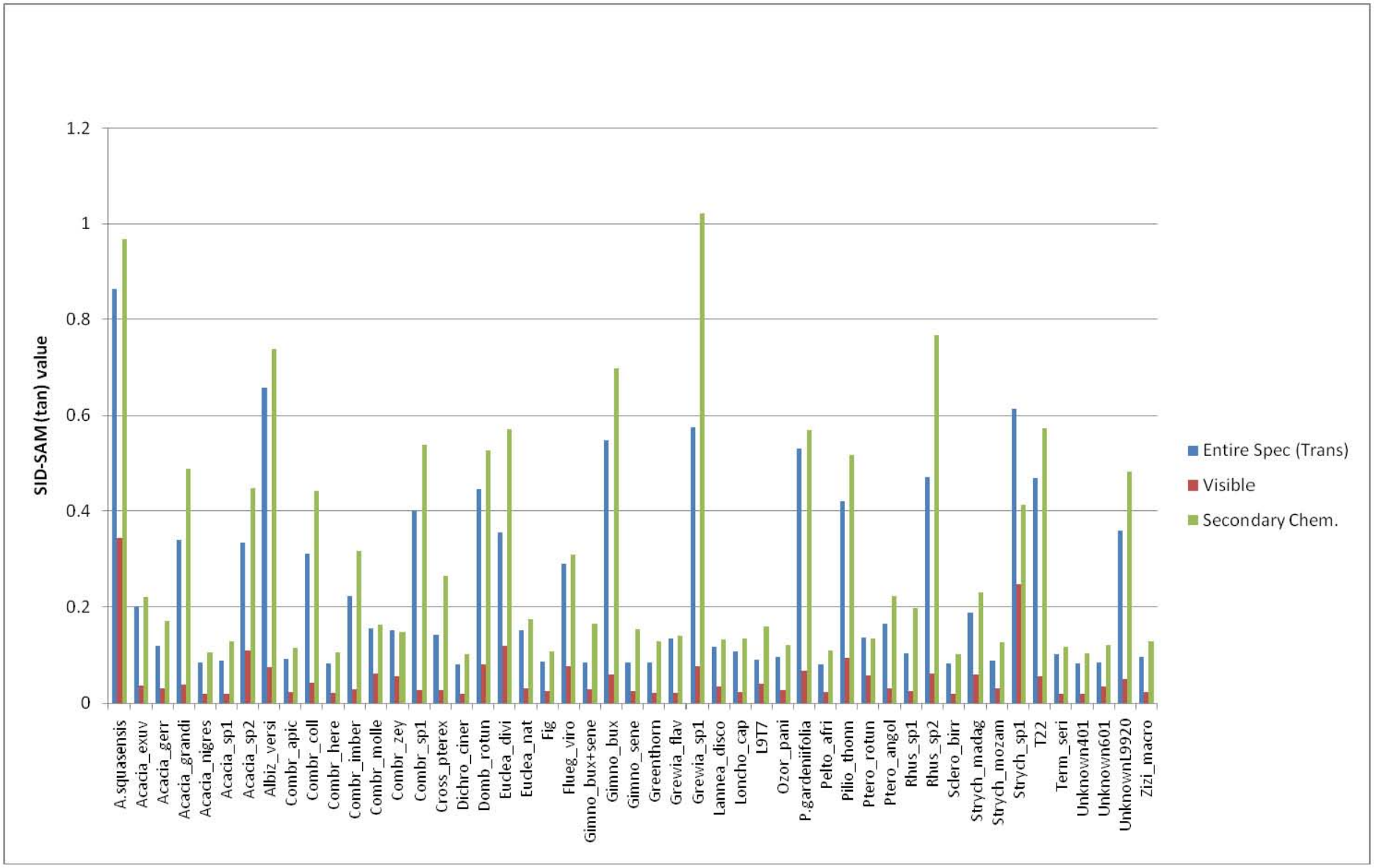


Figure 2.6: SID-SAM (tan) values for different species across 3 different spectral scenarios

2.3.1 SAM RESULTS

From the overall accumulated SAM values across all 14 spectral scenarios, displayed in figure 2.1, it is evident that the transformed (continuum removed) entire spectrum was the spectral scenarios which yielded the highest overall SAM value (257.94). This result was followed very closely by the primary and secondary plant chemical wavelength combinations which yielded an overall SAM value of 250.99. The spectral scenarios which yielded the lowest ability to discriminate by the way of SAM were transformed blue, transformed green and transformed near infrared regions with total SAM values of 18.49, 12.73 and 21.64 respectively. These broad trends above were reinforced by the species specific results displayed in figure 2.2.

Three major spectral scenarios displayed in figure 2.2 – entire spectrum (transformed), primary and secondary plant chemical wavelengths and the visible region – were chosen to represent the entire range of results obtained from the SAM matrices and to avoid unnecessary cluttering of the graph. Majority of the species in the sample yielded the highest individual SAM values under the transformed entire spectrum scenario. The species which yielded the highest SAM values under this particular spectral scenario were *A. squasensis* (9.02), *Albizia versicolor* (8.27), *Combretum sp* (6.94), *Dombeya rotundifolia* (7.07), *Gimnosporia buxifolia* (7.96), *P. gardeniifolia* (7.62), *Piliostigma thonningi* (6.88) and *Strychnos sp* (8.40). The remaining species yielded lower SAM values but still showed the greatest spectral discriminatory ability in the transformed entire spectrum. The few exception species were *Grewia*, *Acacia grandicomuta*, *Gimnosporia senegalensis* and *Rhus* species 1 and 2 which yielded higher spectral discrimination in the primary and secondary plant chemical wavelengths with 9.49, 6.63, 4.24, 4.63 and 8.33 SAM values respectively. In figure 2.2, the visible region SAM results were the worst performing spectral scenario with consistently lower species specific SAM values in comparison to the entire transformed spectrum and the plant chemical wavelengths. The bush encroaching species; *Combretum apiculatum* (3.81), *Dichrostachys cinerea* (3.72), *Pterocarpus rotundifolius* (4.42), *Strychnos madagascariensis* (5.17), *Strychnos mozambiquensis* (4.17), *Terminalia sericea* (4.00), *Acacia gerrardii* (4.42) and *Acacia exuviales* (5.09); did not yield significantly high SAM values.

2.3.2 SID RESULTS

The SID results, in figure 2.3, differed from the SAM results in that the primary and secondary plant chemical wavelengths yielded the highest overall accumulated SID values with a total measure of 79.33. This spectral scenario also displayed a noticeable margin over the entire spectrum results (both transformed and non-transformed) which yielded overall SID values less than 60. This spectral scenario showed great dominance over the remaining spectral scenarios which only achieved overall SID values no greater than 40. Similarly to the SAM results, the transformed green, transformed blue and transformed Near-infrared spectral regions yielded the poorest results with overall SID values close to 0. In contrast to the transformed entire spectrum SAM result; it was clear that the continuum removed transformation had no effect at all in improving the SID species discrimination results.

The results in figure 2.4 echoed the trends observed in figure 2.3 with the primary and secondary plant chemical wavelengths spectral scenario being the appropriate region for species discrimination as the obtained SID values were consistently higher than the other scenarios at the individual species level. The spectrally distinct species (high SID values) were similar to the ones described in the SAM results. The species which yielded very low SID values (less than 1) and thus a higher degree of spectral similarity for all the spectral scenarios were *Combretum apiculatum*, *Combretum hereroense*, *Acacia nigrescens*, *Dichrostachys cinerea*, *Fig*, *Lonchocarpus capassa*, *Peltophorum africanum*, *Terminalia sericea* and *Sclerocarya birrea*. Similarly to the SAM results and the SID trend above, it was clear that the bush encroaching species, once again, yielded insignificant similarity measure values.

2.3.3 SID-SAM RESULTS

The SID-SAM (tan) results, in figures 2.5 and 2.6, yielded trends that were fairly similar to the SID results. The SID-SAM (tan) results were used in this study, rather than the SID-SAM (sin), as slightly higher values were observed. From figure 2.5, it was evident that the primary and secondary plant chemical wavelengths spectral scenario yielded the highest accumulated SID-SAM mixed value of 14.92 which was significantly more dominant than the remaining spectral scenarios. The transformed green, transformed blue and transformed near infrared spectral scenario, once again, yielded the lowest values that were close to 0.

According to figure 2.6, the following species yielded high SID-SAM (tan) values under the chemical wavelength scenario: *A. squasensis* (0.97), *Acacia grandicomuta* (0.49), *Albizia versicolor* (0.74), *Combretum sp1* (0.54), *Dombeya rotundifolius* (0.53), *Euclea divinorum* (0.57), *Gimnosporia buxifolia* (0.70), *Grewia sp1* (1.02), *P. gardeniifolia* (0.57), *Piliostigma thonningii* (0.52) and *Rhus sp2* (0.77). In contrast to the SID results, *Grewia sp1* was found to yield the highest SID-SAM (tan) value and was more spectrally distinct than *A. squasensis*. The SID-SAM (tan) results, from figures 2.5 and 2.6, also differed from the SID results in the fact that the more spectrally distinct species (high similarity measure values) and high performing spectral scenarios yielded higher value ranges while the less significant species and spectral scenarios performed worse and obtained lower value ranges. In agreement with the trend in the SAM and SID results, the bush encroaching species yielded much lower similarity measure values.

2.3.4 PSD RESULTS

To evaluate the performances of the SAM, SID and SID-SAM (tan) discrimination results, the PSD measure was applied to all the matrices results to ascertain the spectral similarity measure (and thus its associated results) which yielded the highest species discriminatory results. The results, from figure 2.7, indicated that the species which showed the highest probability of spectral discrimination under the SID-SAM (tan) measure, were the following 13 tree species (excluding unknown species): *A. squasensis* (2.34), *Albizia versicolor* (0.93), *Dombeya rotundifolia* (0.81), *Euclea Divinorum* (1.07), *Flueggea virosa* (0.78), *Gimnosporia buxifolia* (0.92), *P. gardeniifolia* (0.92), *Piliostigma thonningi* (0.90), *Strychnos sp* (1.92), *Grewia sp* (1.26), *Acacia grandicomuta* (0.74) and *Rhus* species (1.19). The remaining less spectrally distinct species which consisted of approximately 70% of all the species under investigation, including all the bush encroaching species, had lower overall probabilities of spectral discrimination (less than 0.7). However, amongst the less spectrally distinct species, the highest overall probabilities of spectral discrimination were mostly achieved under the performance of the SAM spectral similarity measure. The SID PSD result displayed only intermediate value ranges, throughout, for all the species under investigation.

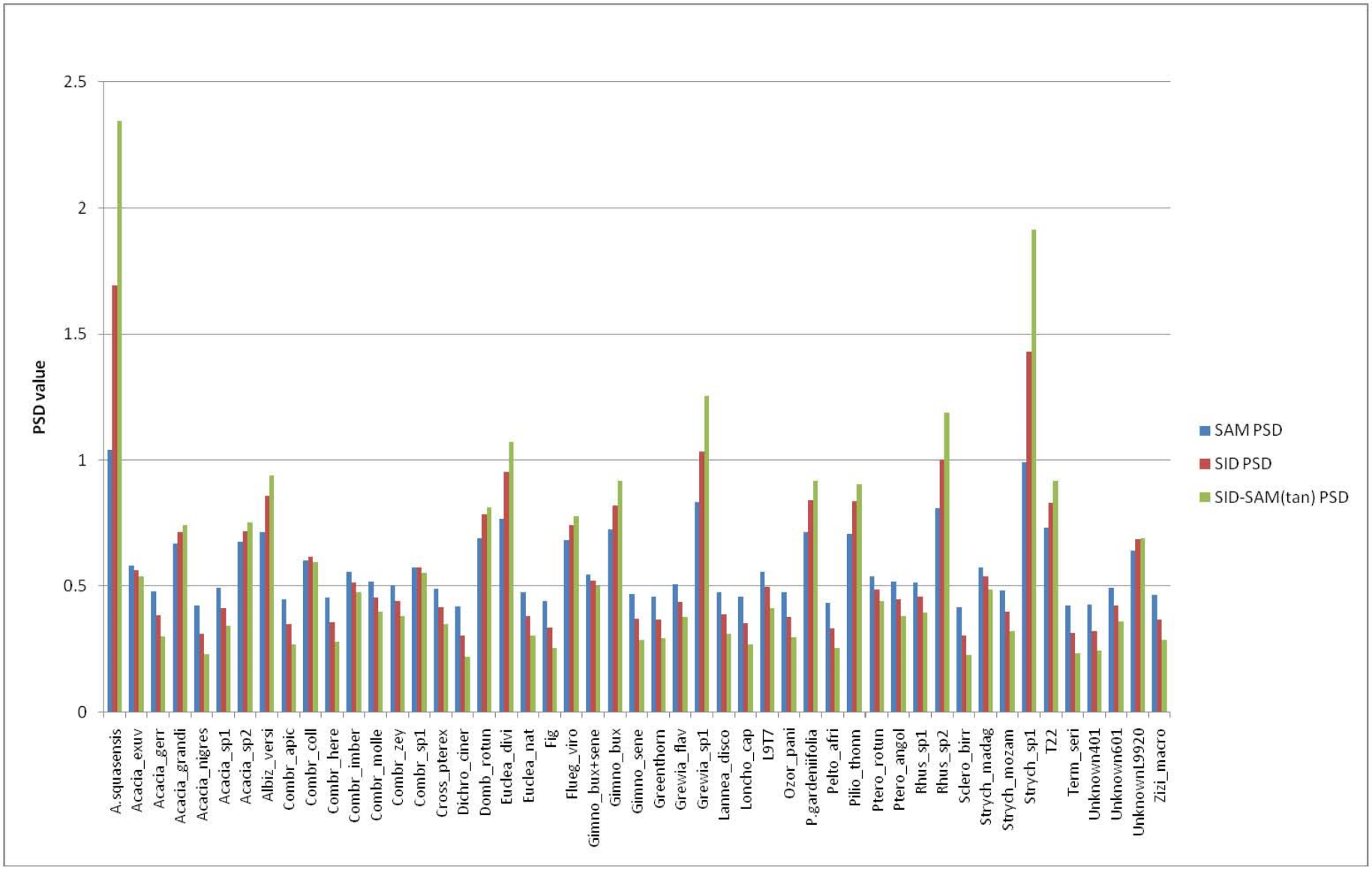


Figure 2.7: Overall PSD results for all the species for SAM, SID & SID-SAM (tan) measures

2.4 DISCUSSION

According to the SAM results, the highest overall SAM values were obtained in the transformed entire spectrum which indicated that the greatest potential for species discrimination was concentrated within this spectral scenario. With the primary and secondary plant chemical wavelength scenario obtaining an overall SAM value that was almost as high, it was evident that the majority of the discriminating potential contained within the entire transformed spectral region may have been concentrated within these primary and secondary plant chemical wavelengths (422.6nm; 432nm; 441.4nm; 450.9nm; 469.7nm; 479.2nm; 488.6nm; 639.9nm; 658.8nm; 932.3nm; 988.7nm & 1044.9nm). From the trends in figure 2.2, as is the case in the SID and SID-SAM (tan) results, it was evident that particular species were more spectrally separable or spectrally distinct than others as they achieved higher SAM, SID or SID-SAM (tan) values across the different spectral scenarios. In the SAM results, certain species such as *Grewia, Acacia grandicomuta, Gimnosporia senegalensis* and *Rhus* species 1 and 2 illustrated a different trend by yielding greater spectral discrimination in the primary and secondary plant chemical wavelengths than the transformed entire spectrum. This trend was possibly due to the fact that the particular species canopies, from which the reference spectra were collected, were in a significantly drier phenological state than the other species and thus the secondary plant chemicals such as carotenoids & xanthophylls were concentrated and more expressive in their spectral profiles. Thus for the SAM measure, the region for conducting species discrimination would lie in the entire transformed spectrum or the secondary plant chemical wavelengths if the reduction of data dimensionality and computational time was considered a major priority.

The SID results illustrated that the primary and secondary plant chemical wavelengths were the preferred spectral scenario for species discrimination. This particular trend was similar to the findings in the SID-SAM (tan) results. This observation was considered plausible as it was supported by the fact that the CAO hyperspectral imagery, from which the endmember reference spectra were extracted for similarity measure analysis, were taken during a dry spell in April/May 2008. The savanna tree foliage was noticeably drier with a greater concentration of secondary plant chemicals in the leaves. This expression of plant chemicals would have subsequently influenced the spectral differences in the absorption and reflectance characteristics in the different tree species. In both the SID and SID-SAM (tan) results, the

transformed spectral scenarios performed poorly and did not in any way improve the discriminating ability of the different savanna tree species. These two measures may have been limited by the fact that they were not able to fully utilise the enhanced absorption features of the reference spectra as was the case in the transformed entire spectrum SAM result. The trends observed in SID-SAM (tan) results (figures 2.5 and 2.6) were expected as the SID-SAM measure typically enhances the similarity between similar spectra and enhances the distinctness between dissimilar spectra. As a result, the perceived discriminatory ability of the primary and secondary plant chemical wavelengths and the few spectrally distinct species may be slightly exaggerated in comparison to the observations in the other similarity measures.

Across all the spectral similarity measure results, the following species were mainly identified as the most spectrally distinguishable species among the total 49 species: *A. squasensis, Albizia versicolor, Dombeya rotundifolia, Euclea Divinorum, Flueggea virosa, Gimnosporia buxifolia, P. gardeniifolia, Piliostigma thonningi, Strychnos sp, Grewia sp, Acacia grandicomuta, Gimnosporia senegalensis* and *Rhus* species. This spectral distinctness present in these particular species could be attributed to the distinctive spectral absorption and reflectance features found across the spectral profiles of these species. On the other hand, the bush encroaching species were found to be consistently obtaining poor results across the three spectral similarity measures. These encroacher species, thus, were not spectrally distinctive enough to be spectrally separable from one another and from the rest of the savanna tree species. Interestingly, the most spectrally distinct species were *A. squasensis* and *Grewia sp1* while the most spectrally similar species was *Sclerocarya birrea*. The *A. squasensis* and *Grewia sp1* results may have been a biased result as, according to the field data, it was only encountered once or twice during the field visit and was thus considered statistical outliers. The lack of sufficient reference spectra for these species suggested the improbability of these results. The possible reason for *Sclerocarya birrea* (Marula tree) having the lowest spectral separability was due to the fact that majority of its leaves were lost during the dry spell, at the time which the CAO hyperspectral imagery were taken. These woody bare or partially bare tree canopies may have been spectrally confused with other dry tree species which lacked enough leaves for appropriate reference pixel extraction and possible discrimination (field observations, April 2008).

Although the SAM measure was the least adequate for discriminating the 13 spectrally distinct species, according to the PSD results, it performed the best in discriminating the remaining spectrally similar tree species and was, overall, considered more beneficial than the SID and the SID-SAM (tan) measures. This observation starkly contradicted the findings of Sobhan (2007-a) and Du et al., (2004), that favoured the SID-SAM mixed measure, and the Van der Meer (2006), that favoured the SID measure. The SAM measure and its associated results proved effective in spectrally discriminating all the species, especially the bush encroaching species, in the spectral library. Since a large number of the savanna species under study did not show any significantly high degrees of spectral separability, the SID-SAM mixed measure would not have been a wise choice in the overall discrimination of savanna tree species in this study. This is due to the fact that it would have promoted the emergence of outlier results that favoured the few significant species and disadvantaged the remaining lesser significant species. This mixed measure would have succeeded in an environment with very high spectral variability amongst the specimens (tree species or vegetation types) being investigated. For that reason, the mixed SID-SAM measure proved successful in the Sobhan (2007) and Du et al., (2004) studies. The SID measure, which obtained intermediate probabilities both in the spectrally distinct species and the spectrally similar species, was not considered desirable as it could not provide high probabilities of spectral discrimination for any of the species under investigation. Savanna environments are complex with a higher degree of spectral confusion within the different tree and grass elements which would make species discrimination naturally challenging.

2.5 CONCLUSIONS

According to the overall PSD results, the Spectral Angle Mapper (SAM) spectral similarity measure was proven to be most effective for spectrally discriminating all the species in this study's spectral library. Although the spectral separability of most of the species were not significantly high, it could still be feasible. The use of continuum removed transform function on the reference spectra proved to be very valuable with the entire transformed spectrum showing the highest spectral discriminatory ability in the SAM results but proved ineffective for the SID and SID-SAM (tan) results. It could thus be concluded that the entire transformed spectrum would be the optimal choice for species discrimination. The bulk of this discriminatory potential in the entire transformed spectrum was found to be concentrated in the primary and secondary plant chemical spectral wavelengths (422.6nm; 432nm;

441.4nm; 450.9nm; 469.7nm; 479.2nm; 488.6nm; 639.9nm; 658.8nm; 932.3nm; 988.7nm & 1044.9nm) where chlorophylls -a and –b, carotenoids and xanthophylls were most expressive. With the secondary plant chemical spectral regions being marginally close to the entire transformed spectrum results, in terms of overall accumulated SAM values, and with it being the most effective spectral scenario in the SID and SID-SAM (tan) results, it is clear that these secondary plant chemical spectral absorption wavelengths aid in savanna tree species discrimination. The SAM, SID and SID-SAM (tan) results, however, illustrated that the bush encroaching species in this study (*Combretum apiculatum, Dichrostachys cinerea, Pterocarpus rotundifolio, Strychnos madagascariensis, Strychnos mozambiquensis, Terminalia Sericea, Acacia gerrardii and Acacia exuviales*) yielded fairly low spectral discriminatory values as their spectral properties were considered to be fairly similar to surrounding species. This would make species separation and subsequent mapping very challenging. Ultimately, the SAM measure would be the most effective to utilise whatever spectral differences that may exist to spectrally differentiate the bush encroaching species from the other non-encroaching savanna tree species. The methodology introduced in the next chapter, however, would need to take this poor spectral distinguish-ability into account for the mapping process.

2.6 RECOMMENDATIONS FOR SPECIES MAPPING PROCEDURE

As a parallel study to this thesis chapter, Cho et al (2010) investigated the implementation of a multiple endmember SAM approach to improve the spectral discrimination of savanna tree species. In the study, a mathematical procedure, known as the Band Add-on procedure, was utilised to select and identify the most appropriate bands for species discrimination. This algorithm selects the bands that maximises inter-species SAM and starts off by selecting the two bands which have the highest average SAM, among all pair-wise combinations. It then adds the next consecutive important bands until no significant bands are left (Cho et al, 2010). A total of 31 bands (which occupied a combination of blue, red edge, near-infrared and chemical spectral bands) out of the 72 bands were found to be the most spectrally significant. Since the optimum number of the significant, species discriminatory bands within the entire 72 band spectral range was identified in the Cho et al (2010) study, it was thus recommended that these 31 bands be used for the mapping process rather than the significant but broad spectral scenarios or regions identified in this thesis chapter.

CHAPTER 3: THE MAPPING OF SAVANNA BUSH ENCROACHING SPECIES USING MULTIPLE ENDMEMBER SPECTRAL ANGLE MAPPER IN A NOVEL SPECTRAL AND STRUCTURAL DECISION TREE APPROACH

OVERVIEW

The detection and mapping of tree species in the savanna ecosystem can provide numerous benefits for the managerial authorities of different land use practices and backgrounds. For rangeland monitoring, the accurate mapping of the spatial distribution of economically viable trees could prove to be enormously beneficial for the local populace as these tree species serve as vital sources for food production and fuel wood requirements. For protected and conserved land monitoring, the possible mapping of problematic alien invasive and bush encroaching species could be invaluable for the development of effective ecosystem management strategies to curb related land degradation problems. Since no effective remote sensing technique or methodology was currently available to address this challenge of savanna tree and bush encroaching species mapping, a new novel Decision Tree multiple endmember SAM classification approach was introduced in this chapter study. 4 Spectral vegetation indices and a minimum vegetation height structural parameter were incorporated for the creation of the original decision tree design which was later subjected to 5 different strategies in an attempt to improve the overall SAM classification accuracies. A 2 end-node decision tree design, based on minimum vegetation height thresholding only, yielded the highest overall SAM classification accuracy of 62.2% which outperformed the traditional multiple endmember SAM classification method. This final decision tree design was implemented to map 23 savanna tree species across the Sabi Sands Wildetuin and Bushbuckridge study region. From the classified end node images, used to create the 23 species map, 6 bush encroaching species were extracted for individual mapping. Both the 23 savanna tree species map and the 6 bush encroaching species maps were used to shed some light on the distribution patterns of these species across different land uses, landscape features and precipitation levels. From the bush encroaching species maps, Dichrostachys cinerea was the most prolific bush encroaching species with a dense overall distribution while Combretum apiculatum was the least abundant. These map products and the observed trends would prove extremely useful in assisting the different components of management (rangeland and protected area) with effective decision-making in the study region.

3.1 INTRODUCTION

3.1.1 WHY MAP SAVANNA TREE SPECIES?

Functional plant type and especially plant species level identification and detection has always been a topic of enormous interest shared by scientific (remote sensors and ecologists) and managerial spheres of society (National Park and communal rangeland management). The detection and mapping of tree species in the savanna ecosystem can provide numerous benefits for the managerial authorities such as the accurate mapping of the spatial distribution of economically viable trees which are a key source of food production and fuel wood for the local populace and communities. The Marula Tree (*Sclerocarya birrea*), for example, plays an important role as non-timber forest products (NTFPs) for the local community enterprises in the communal rangelands who utilise the Marula fruit for beer brewing in cultural and especially trading activities (Shackleton and Shackleton, 2003). These types of enterprises could serve as a safety net for poverty alleviation. Species such as *Terminalia sericea* are primarily utilised as fuel wood to satisfy the energy requirements of the local communities. Other species such as *Pterocarpus angolensis*, *Dalbergia melanoxylon* and *Spirostachys africana* are prominently utilised in woodcraft industry (Shackleton and Shackleton, 2003). These economically viable tree species can thus be monitored for sustainable harvesting and can ultimately be protected against overutilization and degradation.

On the other end of the spectrum, plant species detection and mapping can help ecologists and managers to monitor problematic alien invasive species and bush encroaching species. According to the annual reports of the Kruger National Park history, the region was constantly plagued by problematic species such as *Dichrostachys spp*.; *Pterocarpus rotundifolius*; *Strychnos madagascariensis*; *Maytenus senegalensis*; *Dalbergia melanoxylon* and *Commiphora spp* which were identified by the park management as hardy bush encroaching species (Joubert, 2007). This species mapping and monitoring study would greatly help in improving the natural quality of the heavily impacted and sensitive environments. Since a mapping method currently did not exist to decompose the extremely heterogeneous and complex savanna environments at the individual tree species level, numerous available techniques ranging from pixel-based classification to spectral unmixing had to be critiqued in order to ascertain the most suitable approach for this task. This

exercise would thus serve as a much needed step forward into filling the gap in the remote sensing literature.

3.1.2 AVAILABLE TECHNIQUES FOR TREE SPECIES MAPPING, THEIR LIMITATIONS AND THE WAY FORWARD

Pixel-based classification has always been the mainstay of many mapping studies (both past and present) such as Shafri et al (2007), Yingchun et al (2006) and Kim (1996), just to name a few. Classification is a quantitative method of digital image data interpretation in which the spectral attributes of each pixel, in the image, are individually examined and each pixel is given a label identifying it as belonging to a particular class of pixels of interest (Richards and Jia, 2005). The spectral response patterns of different material types present in the image are analysed, during the classification process, by the use of classifiers. These classifiers can either be described as being parametric or non-parametric. Parametric classifiers are based upon statistical assumptions which include the multivariate normal distribution within spectra classes while non-parametric classifiers make no assumptions about the class distributions and are not based upon class statistics such as the mean and covariance (Joseph, 2005 & Schowengerdt, 1983). Most traditional pixel-based classifiers, such as Minimum-Distance-To-Means (MINDIST) and Maximum Likelihood (MAXLIKE), are parametric in nature. Although these parametric classifiers have been widely used in the remote sensing literature, their assumptions do not fit all applications and are difficult to be implemented in complex landscapes where classes can have high variability (Hansen et al, 1996 cited in Joseph, 2005). Also, the emergence of increased dataset dimensionality, due to the advent of hyperspectral imagery, has pushed the limits of these traditional parametric classifiers with varied classification success (ERDAS Inc, 1999).

Due to the inflexible statistical assumptions of most parametric classifiers, they tend to be affected by the 'Hughes Phenomenon', especially in the case of MAXLIKE, which arises in high dimensionality data when the training dataset size is not large enough to adequately estimate the covariance matrices (Cartijo & De la Blanca, 1996). With the 'Hughes Phenomenon', in a fixed training sample size, the classification accuracy first grows and then declines as the number of spectral bands increases (Boschetti et al, 2007). In hyperspectral classification studies, acquiring the sufficient number of training data that exceeds the total number of spectral bands, required for the MAXLIKE classifier, is an impractical and time

consuming feat especially in highly, spectrally variable environments. These limitations resulted in the introduction of more flexible, non-parametric classifiers such as K- Nearest Neighbour (KNN), Artifical Neural Network (ANN), Decision Trees and Spectral Angle Mapper.

In a study by Shafri et al (2007), these non-parametric classifiers were tested against MAXLIKE, a popular parametric classifier, in a hyperspectral image analysis of a relatively homogeneous tropical forest. According to the overall classification accuracies, the MAXLIKE classifier yielded optimal results with SAM yielding the poorest results. In the study by Kim (1996), however, MAXLIKE was evaluated against SAM in a winter cover classification scenario with SAM yielding the most accurate classification results. This study illustrated that the utilisation of SAM, for vegetation classification, reduced the effects of differential illumination conditions (influenced by variable terrain) and low solar elevation angles typical in the remote sensing of more heterogeneous environments. The effectiveness of this SAM classifier in hyperspectral classification studies was also demonstrated in the Yingchun et al (2006) study which involved the classification of different mangrove types. The results from chapter 2 also supported the importance of SAM, especially in the light of savanna tree species spectral discrimination. The SAM classifier was thus considered as a favourable component for this study's mapping methodology as it was insensitive to differences in illumination, shadowing and albedo (and the 'Hughes Phenomenon'), which are problems associated with airborne hyperspectral imagery (Cho et al, 2010). Other non-parametric classifiers such as KNN and ANN were not considered. ANN was too computer intensive and time consuming due to the level of complexity and customisation that was required. It was also difficult to determine the optimal K value for the KNN classifier (Joseph, 2005).

A vital study by Cho et al (2010) illustrated the importance of multiple endmember usage in the classification process. In the study, we investigated the use of a multiple endmember SAM approach to improve the discrimination of savanna tree species. When discriminating 10 common savanna tree species, it was determined that multiple endmember SAM approach yielded much higher accuracies than the traditional SAM approach which utilised a single mean endmember per species (Cho et al, 2010). The use of a single mean endmember per species in the classification process assumes a unique spectral identity for each species which does not account for the natural intra-species spectral variability expected in reality (Cho et

al, 2010). According to Cho et al (2010) the effect of this intra-species spectral variability, which is relatively high in savanna environments (Cho et al, 2009), was minimized by adopting a multiple endmember approach. This was because multiple endmember spectra serve as multiple identifiers for each particular species during the classification process. It is important to note, however, that the multiple endmembers for each species should accurately represent the intra-species variation that exists in the population under investigation (Cho et al, 2010).

With the expansion of the remote sensing knowledge base and associated technology, modern land cover quantitative studies, vegetation classification and research have shifted from the use of discrete class (pixel-based) classification to more continuous methods such as sub-pixel spectral unmixing (Chambers et al, 2007). Spectral unmixing or spectral mixture analysis (SMA) is defined as a method which determines the fraction of the reflected spectral signal that contains a target (endmember) or targets of interest (Chambers et al, 2007 & Jafari et al, 2006). These target(s) of interest to be detected from the imagery can range from different ground cover components (such as photosynthetic vegetation, non-photosynthetic vegetation and bare soil in Asner & Heidebrecht (2002)) to vegetation types (such as river plant functional types in Kooistra et al (In press)) to vegetation species (Mediterranean tree and shrub species richness in Sobhan (2007 – b)). According to Sobhan (2007 – b), spectral unmixing techniques are essential in classifying imagery where the individual vegetation communities in a landscape (e.g. canopies) are smaller than the pixel resolution of the sensor. For this thesis study, the 1.12 m pixel resolution of the CAO imagery was sufficient to spectrally cover and represent most of the tree canopies, for the purposes of this study, without much background spectral confusion from the understorey canopy.

When considering the most optimum technique for mapping tree species in a complex heterogeneous savanna environment, it was evident that a single technique on its own was not suitable. To overcome the inter-species and high intra-species spectral variability, typical of savanna environments (Cho et al, 2009 and Cho et al, 2010), and the limitations of the previously explored techniques above, a holistic Decision Tree approach was suggested. A decision tree is made up of a series of nodes that are created via rule sets of various parameter thresholds, which are used to subset the dataset under investigation into manageable groups until an end node for classification or a defined class is reached. This process involves the automatic selection of features that carry the maximum information while rejecting the

remaining features (Affendi et al, 2009). Such a decision tree approach, together with the use of a decision tree classifier (e.g. the binary Classification and Regression Tree or CART), had been successfully implemented in the Affendi et al (2009) study for the mapping of tropical forest canopies in Malaysia. This holistic approach also proved vital in the identification and mapping of invasive aquatic vegetation in a hyperspectral remote sensing analysis of the California Delta by Hestir et al (2008). In the Hestir et al (2008) study, a decision tree approach was also implemented but the design and classification procedure were augmented by the use of SMA, SAM, band indices and continuum removal parameter rule sets to overcome mixed pixels which contained similar spectral constituents. Decision trees excel in that they are highly repeatable for a given scenario and increases computational efficiency while overcoming data dimensionality problems (Affendi et al, 2009). Decision trees are strictly non-parametric, and thus not affected by the 'Hughes Phenomenon', and are flexible enough to allow the integration of other classifiers and mapping techniques (such as SAM). Decision trees are also insensitive to missing data values, are capable of making use of both numerical and categorical input data simultaneously and can deal with non-linear relations (Shafri et al, 2007). The success of the decision tree approach, however, strongly depends on large amounts of reference data, which are representative of the different classes and their spectral variability, and the effectiveness of the rule sets defined by the user i.e. the overall decision tree design (Lieng et al, In. press).

Apart from the purely spectral approaches, mentioned above, the incorporation of the structural parameter of vegetation height in the various classification approaches proved to be a valuable addition. Active structural sensors, such as LiDAR and RADAR, produce data products such as Digital Elevation Models (DEM) and Surface Elevation Models (SEM) which provide useful information about the vertical location and horizontal distribution of tree canopy elements (e.g. vegetation height, biomass, cover, density and canopy volume). The inclusion of LiDAR derived structural data, in classification studies, had benefitted the separation of classes, especially those which shared similar spectral properties (Bork & Su, 2007). Bork & Su (2007) utilised an integrated LiDAR data and multispectral imagery approach to enhance the classification of rangeland vegetation. The LiDAR data was utilised in the mapping process by detecting the differences in vegetation height and then implementing vertical height 'thresholds' for the adequate height separation of the different vegetation communities (Bork & Su, 2007). The classes that could not be adequately classified with the LiDAR data only (e.g. bare soil and grasslands) were subjected to a

MAXLIKE classifier for classification in the multispectral imagery and the results were combined. Geerling et al (In. press) combined image spectroscopy and LiDAR data, by data fusion at the pixel level, to improve the classification of floodplain vegetation types. The layers of image spectroscopy bands and LiDAR pre-processed texture bands were stacked to form the fused images which were classified using the MAXLIKE classifier (Geerling et al, In. press). The inclusion of the LiDAR data component in the spectral classification process significantly improved the overall accuracies than the LiDAR data or spectra alone. Lastly in the invasive tree species detection study by Asner et al (2008 – b), LiDAR was used to isolate the sunlit portions of tree crowns of a minimum prescribed height, from hyperspectral imagery, for spectral unmixing. LiDAR derived structural vegetation height was thus seen as an important parameter to be included in this study's methodology as it aided in stratifying the tree canopies to assist the spectral classification of different sized tree species and aided in removing the spectral contamination caused by the understorey and shadow.

From the robustness of decision trees to the effectiveness of SAM as a significant discriminatory measure to the advantages of utilising multiple endmembers in the SAM classification and to the importance of structural vegetation height, a decision tree classification approach, was therefore suggested as an adequate method to integrate the multiple endmember SAM classifier and LiDAR derived tree height variable for classification of the tree species.

3.1.3 AIMS, OBJECTIVES AND RESEARCH QUESTIONS

The main aim of this chapter is to introduce and test a novel decision tree multiple endmember SAM classification technique to map bush encroaching savanna tree species in the Greater Kruger National Park area across different land use types. This technique is novel because it is the first of its kind to integrate multiple endmember spectral classification and structural height thresholding into a single approach to map savanna tree species in complex savanna environments.

The ultimate goal is to describe the general pattern and distribution of the savanna bush encroaching tree species in the Greater Kruger National Park region with special focus on the Bushbuckridge and Sabi Sands Wildetuin areas. Approximately 23 tree species was focused on for classification which differed from the last chapter's 49 species. Once the savanna tree

species have been classified, the bush encroaching species were isolated for mapping purposes. Six main bush encroaching species were focused on which were analysed and discussed in detail – *Acacia gerrardii, Combretum apiculatum, Dalbergia melanoxylon, Dichrostachys cinerea, Pterocarpus rotundifolius* and *Terminalia sericea* (Joubert, 2007).

This thesis chapter sets out to answer these three research questions:

- How effective is this newly proposed Decision Tree multiple endmember SAM classification approach relative to a traditional multiple endmember SAM classification approach? Which decision tree design obtained the highest overall classification accuracy?

- How does the distribution and composition of each of the bush encroaching species differ along the L456 study region, from Bushbuckridge Municipal District to Sabi Sands Wildetuin?

- Which one of the bush encroaching species was the most prolific and which was the least? What are the implications of these bush encroaching species' distributions on the communal rangeland and protected land management practices?

3.2 METHODOLOGY

After much trial and error, the following methodology was put forward as the optimal available approach to achieve the goal of mapping savanna bush encroaching tree species. This chapter's methodology depended on the outcomes of chapter 2 as it pointed out which bands or spectral regions would be used in the classification algorithm which was in turn dependent on the best performing spectral similarity measure (SAM). All these factors also influenced the parameters included in the Decision Tree construction process. It was also important to note that all the data used in this chapter's methodology were mostly extracted from the L456 CAO hyperspectral and LiDAR data. Ground data collected during the May 2010 field campaign, which was also included in the methodological process, will be discussed next.

3.2.1 MAY 2010 FIELD CAMPAIGN AND IMAGE DATA SELECTION

From the original 49 tree species sampled from the previous field trip, the most common tree species, including the bush encroaching species, were narrowed down to 23 tree species and their data libraries enhanced for the mapping process. The remaining species were excluded from the mapping process because they were only sparsely present in the L456 study region.

A more intensive sampling technique was adopted for the May 2010 field campaign. Prior to the field visit, snap shot sub-images of the hyperspectral imagery, covering the study region, were compiled at a resolution in which individual tree canopies were clearly visible. Within these snap shot sub-images, prominent tree canopies were marked with a point shapefile for navigation and identification once in the field. With this sampling technique, the tree canopies were pre-selected for sampling while other species of interest (e.g. bush encroaching species) that were encountered, plus their general distribution, were demarcated on these image snap shots while in the field. The pre-selected tree canopy point shapefiles are displayed in the GIS map (figure 3.1) at the end of this subsection. Spectral and structural information were then extracted from these data sources and combined with the previous April 2009 datasets to create the necessary spectral and structural libraries for the analysis in this chapter. This field sampling technique and the collected data from the 2009 and 2010 field work were only utilized for and limited to the L456 study region image.

The L456 image was chosen over the other images for analysis in this chapter as it was considered the most representative image. This was due to the fact that it preserved the main land use management regimes (protected management of Sabi Sands and the ad hoc communal rangeland management) within a single image. The image also covered a much greater area of land than the other images and the greatest amount of tree species spectral and validation data were also collected from this study region. The incorporation of the collected data into the actual decision tree design, construction and implementation process will be discussed next.

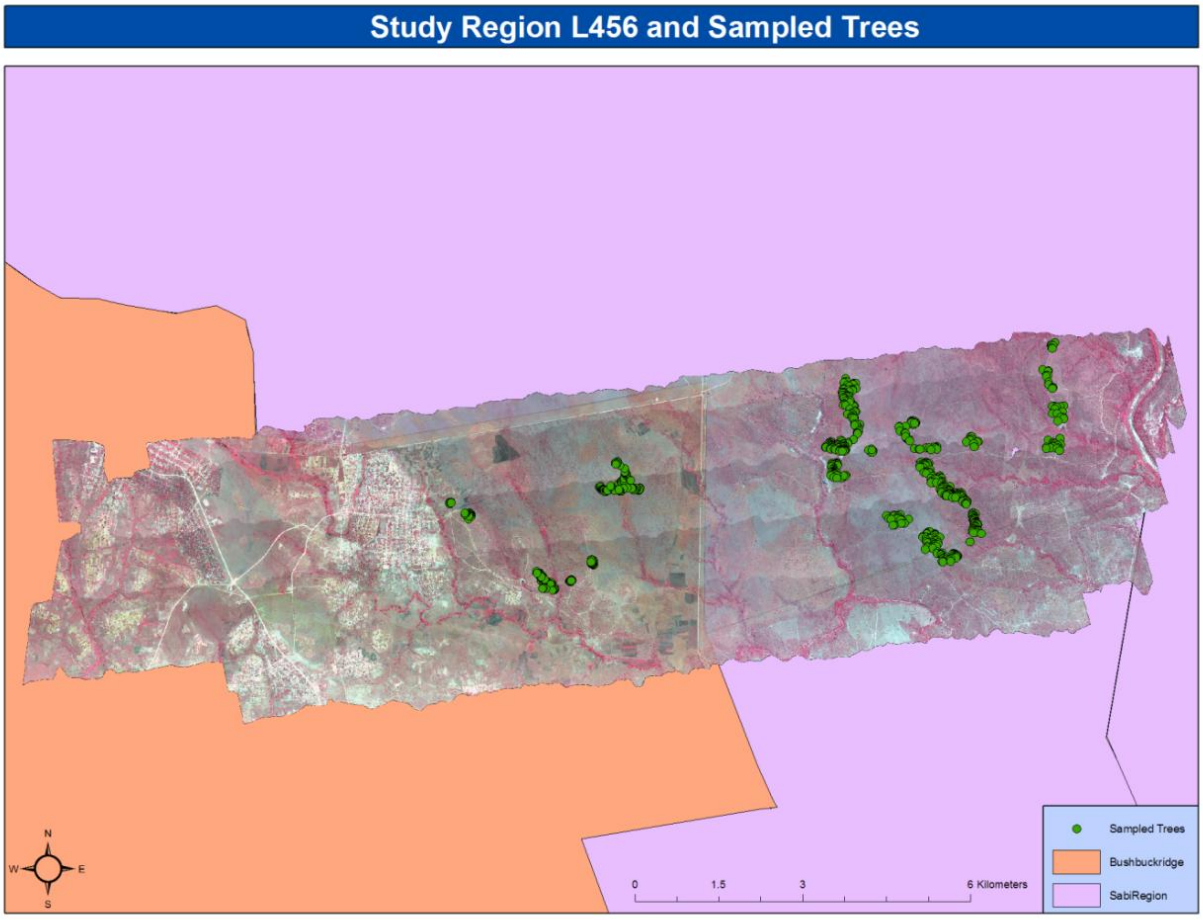


Figure 3.1: Pre-selected and sampled tree canopies of L456 which assisted in the development of the spectral and structural libraries

3.2.2 DECISION TREE DESIGN PROCESS

Data preparation of the various spectral and structural parameters, parameter thresholding for node construction, design validation and classification accuracy assessment were all the vital steps needed for the decision tree design process.

3.2.2.1 DECISION TREE PARAMETER PREPARATION

In this study, a structural tree height library was constructed by overlaying GPS points of different tree species locations over the LiDAR derived vegetation height image. Significant parts of the spectral library constructed in the previous chapter was also utilised, and bolstered by the May 2010 field data, in the construction of this decision tree approach. Just to reiterate, the spectra, for the spectral library, which contained only pure pixels for the

different tree species (23 common species in total) were extracted by using the spectral profile tool (in ENVI 4.7) while the structural parameters (minimum, mean and maximum height) were extracted using the Region of Interest statistics tool in ENVI. Trees that had a minimum height of less than 2 metres were excluded from both the structural tree height and spectral libraries. The reason for this is that the LiDAR functionality was significantly reduced when determining vegetation heights less than 2 metres (Fisher et al, 2009). Each spectrum collected for the spectral library must have its corresponding canopy height measure in the structural library. The overall strategy was to construct, run and validate the decision tree SAM classification in an external Microsoft Excel and Matlab environment prior to the actual mapping process in the ENVI remote sensing software. For this strategy, both the structural and spectral libraries needed to be randomly split into separate training and validation datasets. The training dataset (including both the spectral and structural information) was used solely to construct the decision tree and be used as endmembers in the SAM classification mapping process while the validation dataset was lead through the entire decision tree and the SAM classifier to ascertain the classification accuracy and the overall effectiveness of the decision tree architecture. This will be elaborated upon further. The spectral and structural libraries were split using a random stratified approach which had to be considered based on the intra-species spectral variability that occurred across the L456 study area. Due to a localised but significant rainfall gradient across the L456 study area drier conditions were evident in the Bushbuckridge communal rangelands with wetter conditions in Sabi Sands Wildetuin, which lead to a significant difference in the intra-species spectral properties. This trend is evident in the figures below (figures 3.2 & 3.3) which illustrate the spectral differences between *Sclerocarya birrea* and *Combretum apiculatum* spectra taken from Bushbuckridge and Sabi Sands.

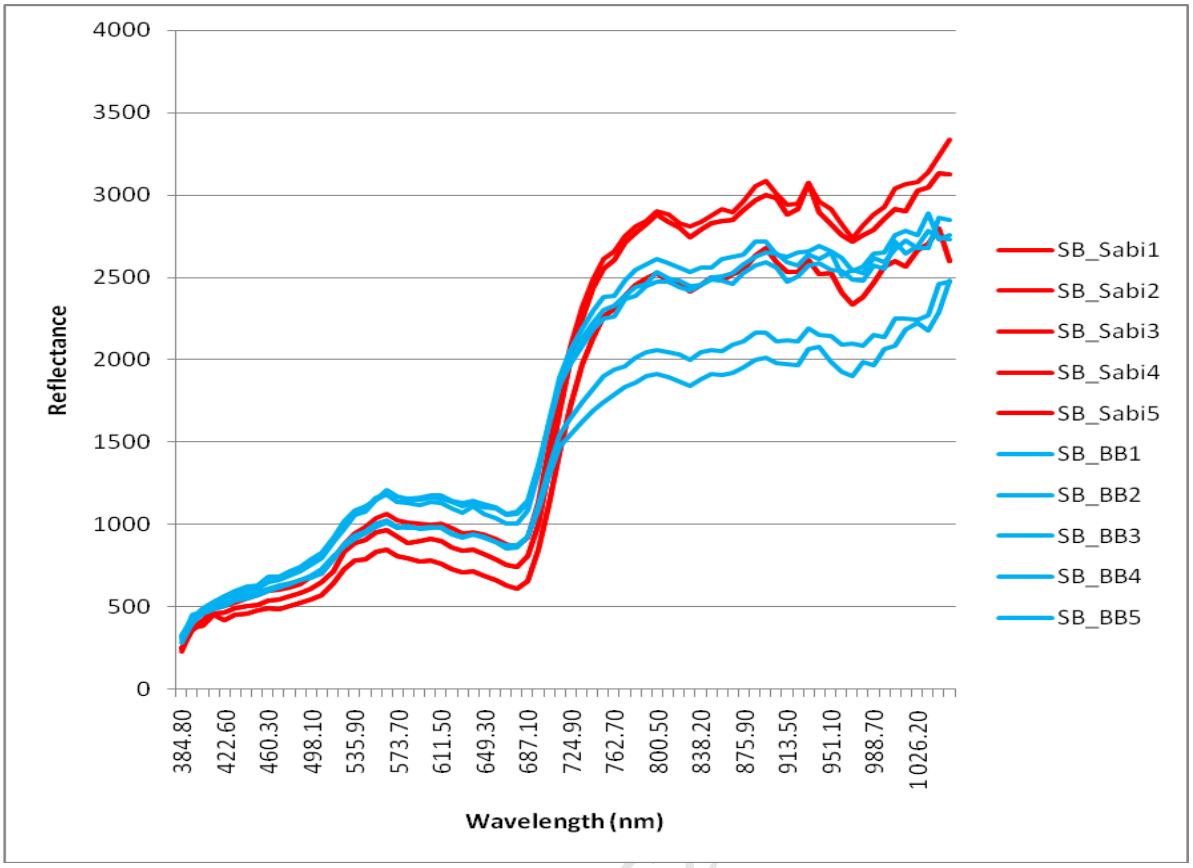


Figure 3.2: ***Sclerocarya birrea*** **reflectance spectra from Sabi Sands and Bushbuckridge**

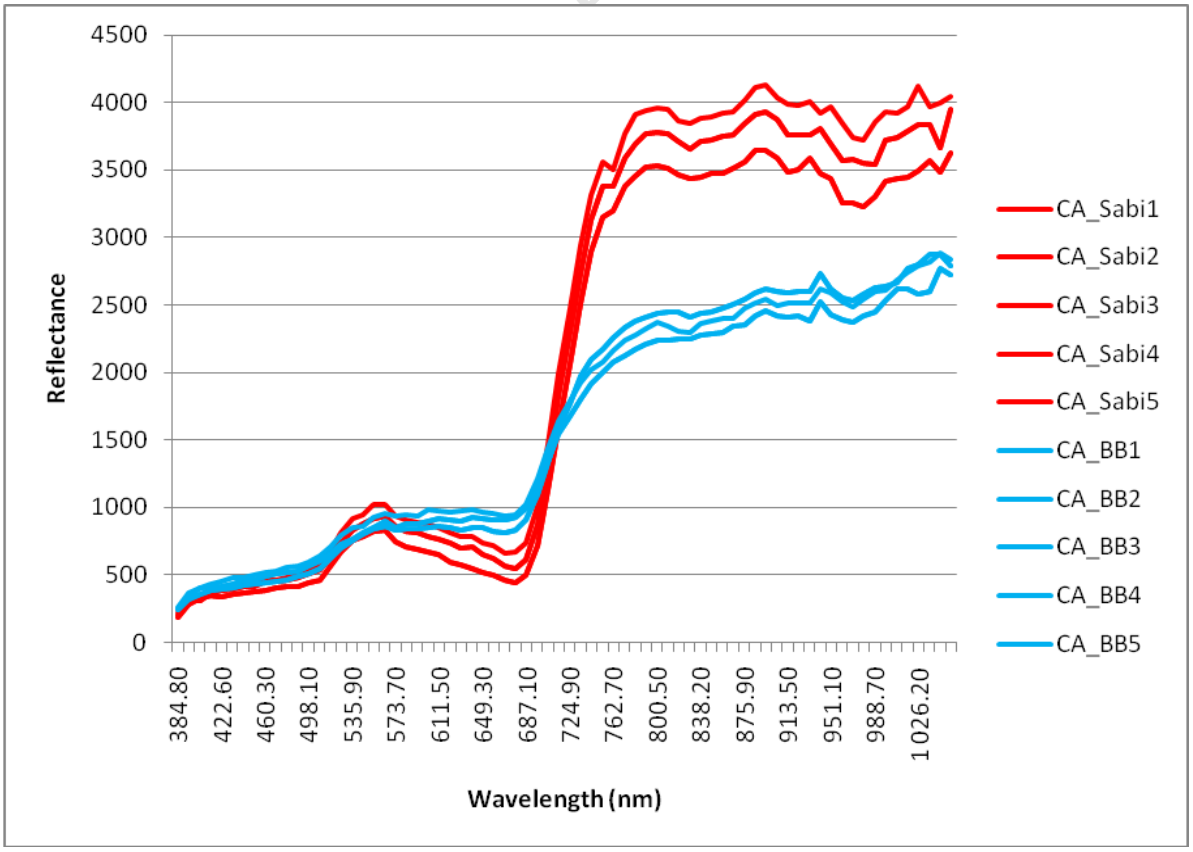


Figure 3.3: ***Combretum apiculatum*** **reflectance spectra from Sabi Sands and Bushbuckridge**

In order to ensure even splitting of the dataset that were representative of this intra-species spectral variability during the random approach, the Bushbuckridge and Sabi Sands spectral and structural data had to be individually split into its own training and validation sub-datasets (hence 'stratified') and then combined to form the master training and validation datasets. For the actual data splitting process, each individual spectrum with its corresponding height measure were numbered numerically for each species which were then split according to the random sequence produced by an atmospheric noise randomness algorithm (www.random.org). After the data splitting process, the total number of pixels sampled, per species, for the training and test datasets was briefly summarized in table found in Appendix 3B.

A combination of spectral parameters and a structural measure were considered as the variables used to build up the decision tree. The spectra from the training spectral library were applied to four vegetation indices: Carotenoid Reflectance Index (CRI), Photochemical Reflectance Index (PRI), Normalized Difference Vegetation Index (NDVI) and a Red Edge NDVI (RE). These four vegetation indices were chosen as they made effective use of the primary and secondary plant chemical wavelengths that were found to be the most useful in the first analytical chapter study. The formulae for these vegetation indices are displayed in table 3.1 below. In essence, it was these vegetation indices values and the minimum tree height statistic which were used as the main parameters to make the necessary decisions in the decision tree's architectural design. The minimum tree height statistic was preferred over the mean and maximum tree heights as the minimum height values would encompass the smallest tree in a particular species' population and would thus allow the mapping of all the possible trees of each species. It was this minimum tree height dataset of the different tree species that was subjected to the thresholding process. The next step in the decision tree construction involved parameter thresholding of the available training data.

Table 3.1: The four Vegetation Indices used and their associated formulae and references

Vegetation Indices	*λ is in nm*	*References*
Carotenoid Reflectance Index (CRI)	λ800(1/λ520 - 1/λ550)	Gitelson et al (2002)
Photochemical Reflectance Index (PRI)	(λ531 - λ570)/(λ531 + λ570)	Gamon et al (1992)
Normalized Difference Vegetation Index (NDVI)	(λ800 – λ678)/(λ800.5 + λ678)	Rouse et al (1973)
Red Edge NDVI (RE)	(λ750 - λ705)/(λ750 + λ705)	Gitelson et al (1994)

3.2.2.2 DECISION TREE NODE CONSTRUCTION

Multiple Variable Histograms for the various indices and the minimum tree height parameter were created in Statistica 6 to visualise the distribution and predicted range of these variables for each species as a whole. From the study by Pu et al (2008), the formulae: "Mean value – 1 Standard deviation; Mean value + 1 Standard deviation" were applied to the structural and vegetation indices values to calculate a relative threshold range [a lower (Lt) and upper (Ht) threshold value] for each parameter and for each species. From this database of threshold values, decision nodes were created to construct the decision tree in which all 23 species were separated into much smaller, manageable end-node groups which were similar in particular spectral and structural characteristics but could not be distinguished further. It was thus the purpose of the classification algorithm to discriminate them further at an individual species level. A total of 9 end nodes were created to make up the original decision tree design for the L456 study region which is displayed in figure 3.4 below. Each end node contained a list of various species that shared the particular parameter's properties and the properties from the previous nodes in the decision tree branching hierarchy. It was these end node species' spectra that were used as the training multiple endmembers for the SAM classification.

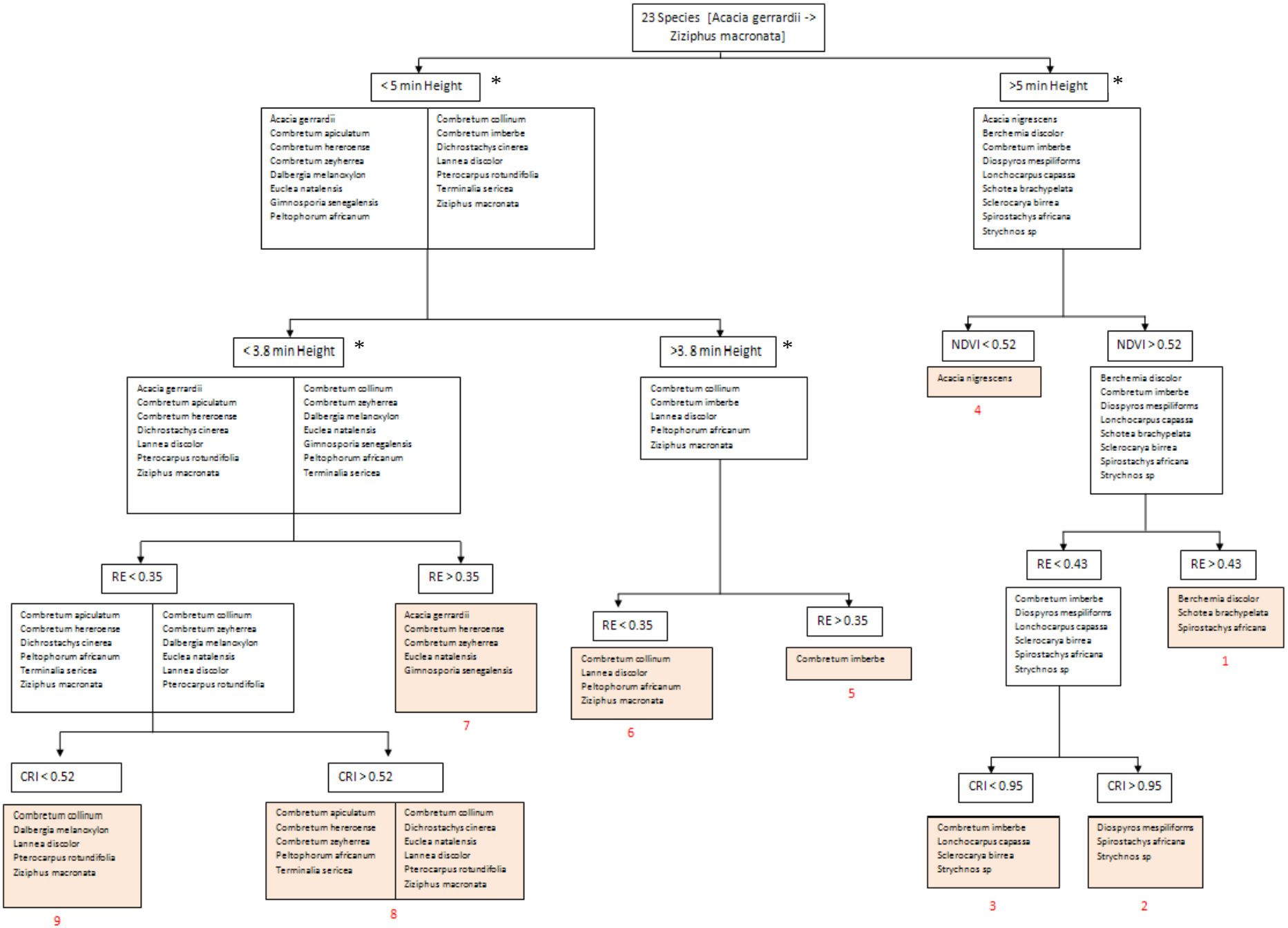


Figure 3.4: Original 9 End Node Decision Tree Design

* All min height threshold values are expressed in metres

3.2.2.3 DECISION TREE VALIDATION AND CLASSIFICATION ACCURACY ASSESSMENT

Once the end nodes were created and the decision tree architecture completed, the validation dataset were then fed through the entire decision tree and were split up while strictly adhering to the minimum height and vegetation indices thresholds at the various nodes. Once the validation data have reached the end nodes, the validation and training spectra for each particular end node were applied to a SAM script classifier (the same script used in the previous chapter analyses) in a MATLAB (version R2009a) environment. The actual script that was used is displayed in the Appendix section 2D of this thesis. The output SAM matrix, for each end node, was then taken to Microsoft Excel 2007 in which the validation or test data were compared against the classified spectra via the "VLOOKUP" function. The minimum SAM value (spectral angle) was used to match the particular test and classified spectra. Confusion matrices were then created using the Pivot table tool in Excel for each node and, more importantly, for the entire decision tree in order to determine the classification accuracy and thus overall effectiveness of the decision tree design. The producer's accuracy, user's accuracy and the overall classification accuracy were the statistics that were obtained from the confusion matrices. The producer's accuracy indicates the percentage of spectra for each species class that have been correctly classified while the user's accuracy indicates the probability that a spectra classified into a given species class actually represents that class on the ground (Baldi and Paruelo, 2008). From the 9 individual node and overall classification accuracy results, various strategies were put forward to improve the overall classification accuracy and reduce error. The intent was to ascertain the decision tree design which yielded the highest classification accuracies prior to the actual mapping process. These five strategies would be explained in the section below.

3.2.3 STRATEGIES TAKEN TO IMPROVE DECISION TREE CLASSIFICATION ACCURACY

The different strategies were summarised in table 3.2 which follows next. After implementing the various strategies, the original 9 end node decision tree (figure 3.4) underwent various design changes in which the resulting end nodes (figure 3.8 on page 82) were subjected to the SAM script classifier and the classification accuracy assessment process.

Table 3.2: Strategies undertaken to improve decision tree classification accuracy

Strategy	Description
Minor Pruning (4 End nodes)	Merging and removing poor performing end nodes. Streamlining decision tree design. Parameter thresholds were kept but were implemented more leniently to accommodate more species with overlapping parameter properties across multiple end nodes. This would reduce classification error.
Minor Pruning + Selected Training Endmembers	Dataset reduction technique in which representative endmembers were selected, using cluster analysis, instead of using the entire spectral training dataset for each species. Single linkage dendrograms were implemented to identify the key endmembers evident within the main groups or clusters.
Minor Pruning + Selected Bands	Utilising significant selected bands, to subset both the training and validation datasets, during the SAM classification process instead of all 72 spectral bands. The 31 bands (*706nm, 762.7nm, 696.6nm, 668.2nm, 677.7nm, 687.1nm, 715.5nm, 724.9nm, 734.4nm, 743.8nm, 753.3nm, 384.8nm, 394.3nm, 403.7nm, 413.1nm, 422.6nm, 913.5nm, 819.3nm, 828.8nm, 838.2nm, 847.6nm, 857.0nm, 866.5nm, 875.9nm, 885.3nm, 894.7nm, 904.1nm, 1016.8nm, 922.9nm, 932.3nm and 941.7nm*) were chosen from the Cho et al (2010) study.
Extreme Pruning (2 End Nodes) + Selected Bands	Limit decision tree design to just the 2 tree height root nodes. All vegetation indices related nodes were removed to evaluate the influence of height alone on the SAM classification and accuracy. Selected bands from the previous strategy was also implemented to improve results.
Traditional Multiple Endmember SAM + Selected Bands	A traditional SAM classification with no influence of a decision tree. Serves as a critical baseline in which all the produced results could be compared to. The optimal decision tree design will be compared to this strategy. The selected bands were also included.

At the end of the analytical process, it was the strategy which yielded the highest overall classification accuracy that would be utilised in the final decision tree design and implemented in the actual mapping process in ENVI 4.7.

3.2.4 BUSH ENCROACHING TREE SPECIES MAPPING

Once the decision tree design and approach was finalised, the next step was to apply this theoretical model, in practice, on the hyperspectral imagery for the creation of savanna bush encroaching species maps. After decision tree implementation and classification, the

classified image products underwent majority class filtering as a form of post-classification analysis. After post-classification analysis, the bush encroaching species were then extracted for individual map representation.

3.2.4.1 DECISION TREE IMPLEMENTATION AND SAM CLASSIFICATION

The decision tree design that yielded the highest overall classification accuracy was replicated, this time, in an ENVI software environment for the actual tree species mapping process. For data preparation, the CAO hyperspectral image of the L456 study region was subset, masked and broken down further according to the nodes of the decision tree to create the end-node images which would be classified. The LiDAR vegetation height image of the study region was used for the height subsetting. For the Vegetation Indices, the NDVI, RE NDVI and CRI indices images were created by inputting the necessary formulae (table 3.1) in the Band Math function in ENVI 4.7 and applying them to the original hyperspectral image. These vegetation indices images were then subset according to the predefined thresholds of the decision tree nodes which was, in turn, used to subset the hyperspectral image. Please refer to the methodology of the previous chapter for image subsetting and masking according to a parameter (e.g. height) threshold. At the end of the image subsetting process, the final products were the hyperspectral images that were representative of the end nodes and their parameter threshold properties of the previous nodes in the decision tree architecture. These hyperspectral images were then subjected to the hard pixel SAM classification algorithm which was performed in ENVI 4.7. The classification procedure made use of the training endmembers of the particular species that were found at the individual decision tree end nodes. The use of multiple endmembers for each species in the SAM classification algorithm would significantly increase the processing requirements and computational time but was essential for a viable and accurate result. The various endmember classes for each species, which were mapped by the SAM classifier, were then dissolved into single species classes.

3.2.4.2 POST-CLASSIFICATION IMAGE FILTERING

In the post-classification stage, these end node classified images were subjected to a 3 pixel by 3 pixel Majority Class Filter in which the majority class within a 3X3 pixel window were extracted for representation. This was done to create clearer classified image products by removing most of the ‘salt and pepper’ mixed classification effect. The filter also permitted

the emergence of a dominant species class in most of the larger individual and clumped tree canopies. This filter, which was mostly successful in the taller tree species classes, however, did not have an entirely desirable effect on the smaller tree species and their canopies as some of the small tree species canopies, e.g. *Dichrostachys cinerea* and *Dalbergia melanoxylon*, were eliminated thus leading to a slight underrepresentation of small tree species classes. Although some of the less prominent, small tree canopies were removed by the majority filtering process, the species' main spatial distribution were still retained so it was considered a small sacrifice for the clear representation of the map products. Since the managers of the different land types (communal versus protected) are more concerned with the main distribution of particular tree species, rather than the exact canopy location of these species, the mapped results would still prove very useful for management purposes.

3.2.4.3 EXTRACTING THE BUSH ENCROACHING TREE SPECIES FOR REPRESENTATION

The bush encroaching species, which demonstrated reasonable classification accuracies greater than 50%, were then isolated from the final, classified decision tree end nodes for individual mapping and distribution analysis. During the classification process in ENVI, the classified pixels of the classified end node products were allocated a range of identity numbers which were directly related to the list of numbers used to identify the species specific endmembers used in the SAM classification. Using the 'band threshold to ROI' function, under the Region of Interest tool, the specific range of identity numbers for the particular bush encroaching species were selected and used to spatially subset the end node classified image (<5m min height) to create an image displaying the distribution of that particular bush encroaching species. This process was repeated for the individual bush encroachment species. The bush encroaching species distribution images were then exported to ArcMap 9.3.1, in a GeoTIFF format, with the original hyperspectral image being used as the backdrop image to illustrate the ground and low lying vegetation (<2m) which comprised of the non-classified parts of the image. The entire methodology is summarised in the flow diagram below.

3.2.5 SUMMARISED METHODOLOGY SCHEMA

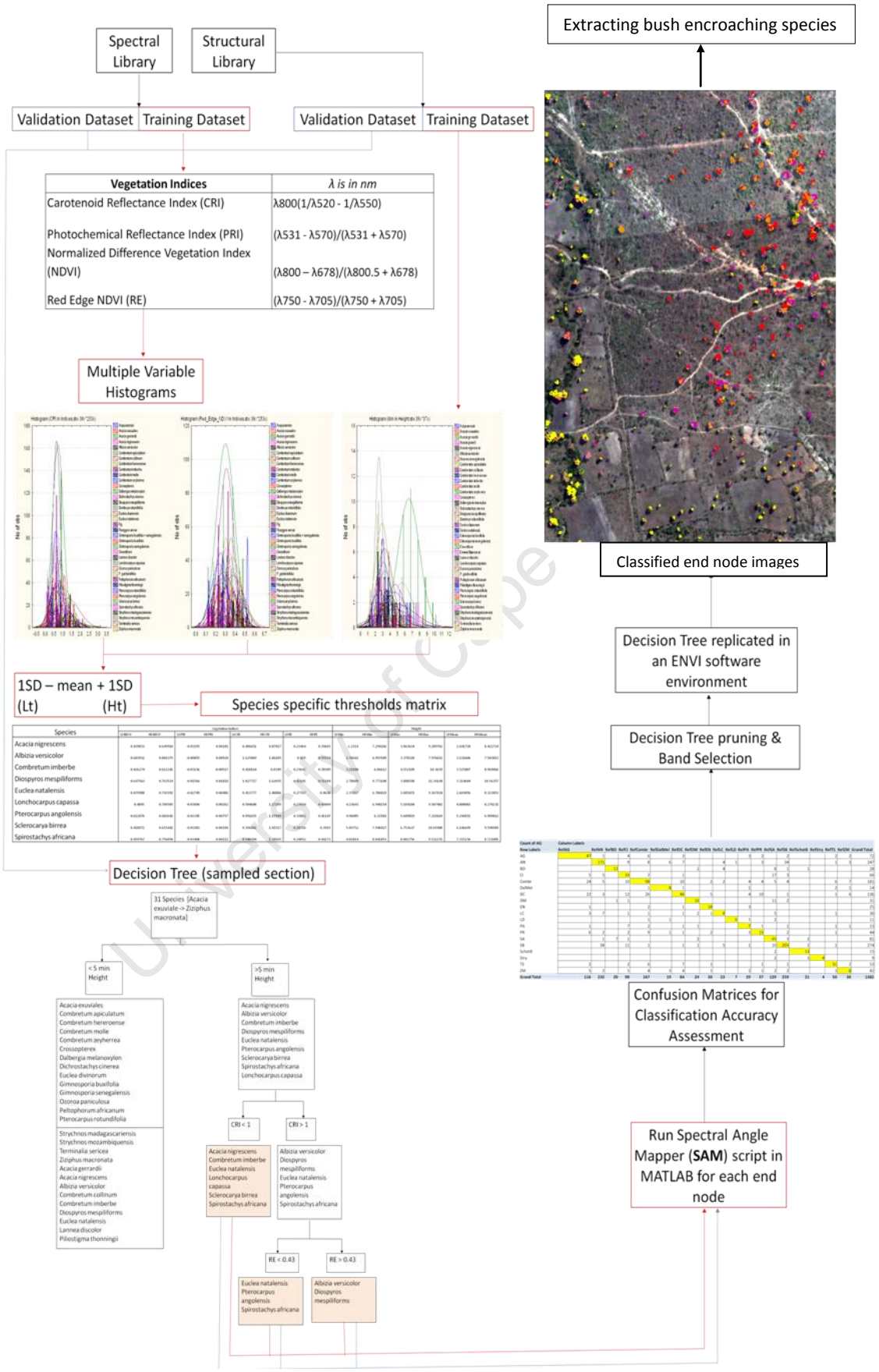


Figure 3.5: Methodology schema for decision tree construction, classification and post-classification

3.3 RESULTS

The results in this section were centred around the discovery of the optimum decision tree design (i.e. the design which yielded the highest overall accuracies). These results included multivariable histogram results (for preliminary variable thresholding), a variable threshold matrix of all 23 different tree species (used to construct the decision tree nodes), and the evolution of the original decision tree design together with associated confusion matrices resulting from the different strategies implemented to improve the overall classification accuracy. A traditional multiple endmember SAM classification was utilised as the baseline result against which the different decision tree designs were compared to. Based on the optimum decision tree design, individual bush encroaching species maps were introduced as the final products for this chapter. In this results section, the 23 tree species each have a designated code with its key available for viewing in Appendix 3A.

3.3.1 MULTIVARIABLE HISTOGRAMS

These histograms were created to display the general distributions of the different variables at a species level and also serves as a 'sneak-peek' to the possible variable thresholds that may exist in the species spectral and structural datasets. From these histograms, the most useful parameters for finding unique value ranges for the individual species could also be ascertained before the exact threshold values were determined for the threshold matrix and decision tree construction. First the structural parameter histogram of minimum vegetation height will be displayed below, which will then be followed by the spectral parameter histograms (NDVI, PRI, CRI and Red Edge NDVI). In these histograms figures, the underlying bar plots illustrate the frequency of the different spectral endmembers used for each species for a particular variable. The normal distribution curves, above these bar plots, describe the general normalized distribution for each species which, in other words, illustrate the expected 'range' for the different species for a particular variable. The amplitude of these curves varies according to amount of spectral endmembers collected for the different tree species which, in turn, differs according to species.

3.3.1.1 MINIMUM TREE HEIGHT HISTOGRAM

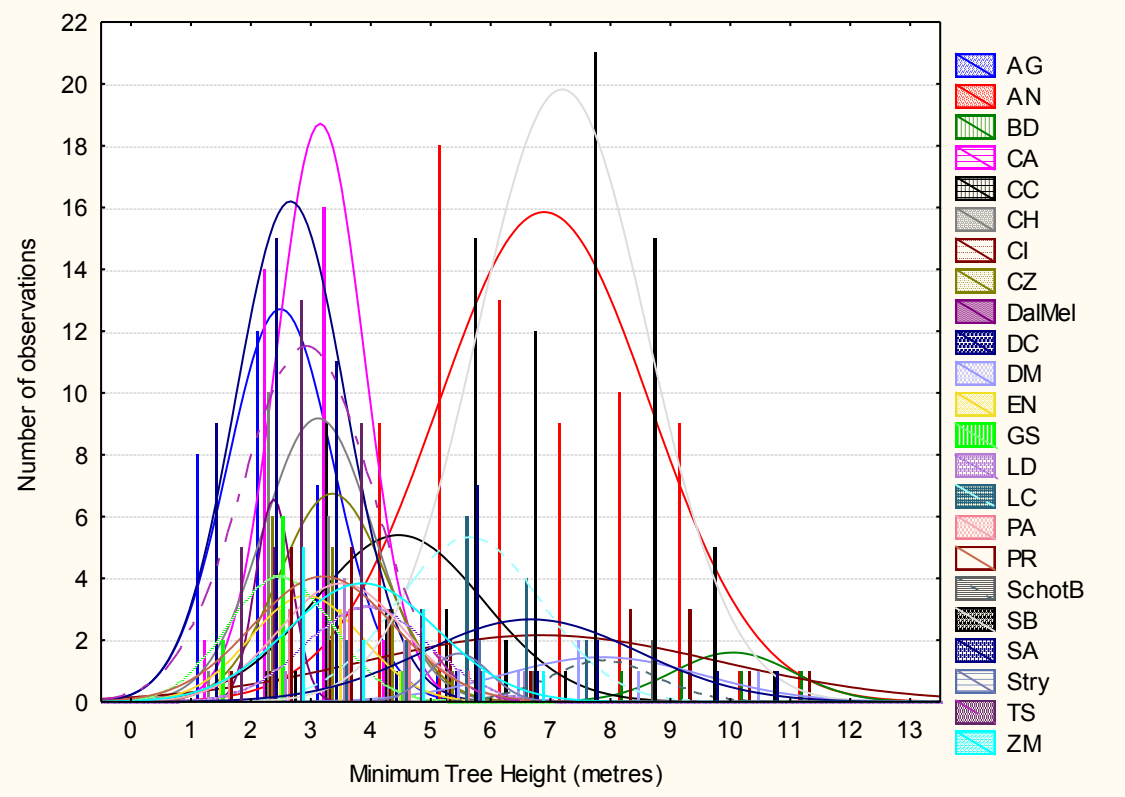


Figure 3.6: Minimum tree height histogram of the 23 savanna tree species

In the histogram above, it was clear that most of the tree species possessed a minimum height distribution that fell within two broad categories – a short tree (~<5m) and a tall tree (~>5m) group. The tree species which fell within the short tree group included species such as *Acacia gerrardii*, *Combretum apiculatum*, *Combretum zeyherrea*, *Dichrostachys cinerea*, *Gimnosporia senegalensis* and *Terminalia sericea* to name a few. On the other hand, the tree species which had a minimum height distribution that fell within the tall tree group include *Acacia nigrescens* and *Sclerocarya birrea* (with slight overlap in the short tree category), *Berchemia discolor*, *Diospyros mespiliformis* and *Schotia brachypelata*. Tree species which had intermediate (~1-9m) minimum height ranges that clearly overlap both height categories include species such as *Combretum collinum*, *Combretum imberbe*, *Lonchocarpus capassa* and *Strychnos species*. *Combretum imberbe* was considered to be the most problematic species as it covered most of the entire sampled minimum vegetation height range for all species, both short and tall (~2-13m). The bush encroaching species fell within the small tree group. The spectral vegetation indices parameters will be addressed next to investigate for possible species specific distinct ranges or distributions.

3.3.1.2 SPECTRAL VEGETATION INDICES HISTOGRAMS

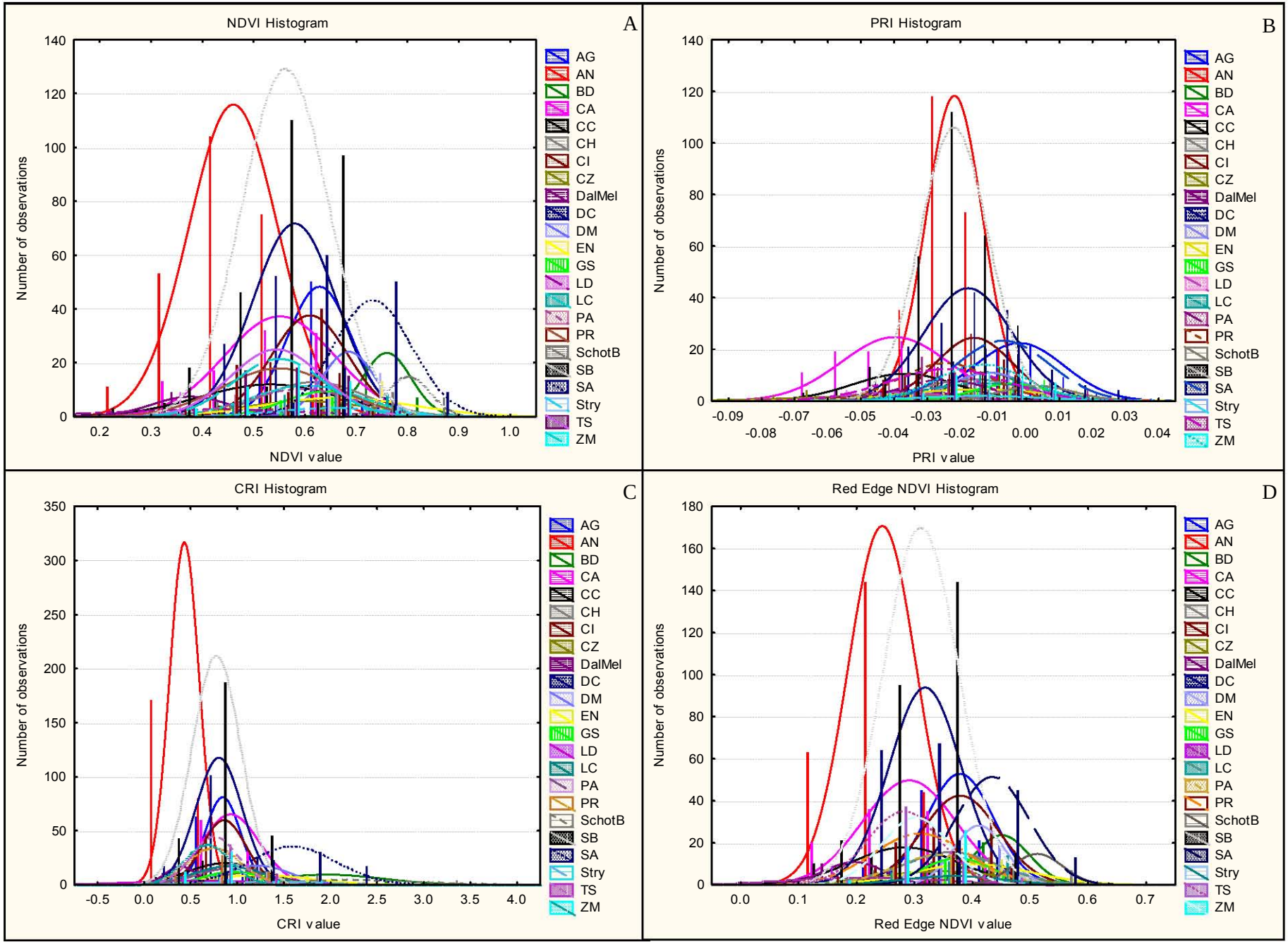


Figure 3.7: Vegetation indices histograms of the 23 savanna tree species – (A) NDVI; (B) PRI; (C) CRI & (D) Red Edge NDVI

In figure 3.7A, the NDVI distribution indicated slightly overlapping ranges with most of the tree species in the spectral library. The lowest NDVI value distribution was observed in *Acacia nigrescens* with the majority of the observations displaying a value around the 0.45 peak value (with a range from ~0.2-0.9). This indicates that most of the *Acacia nigrescens* canopies were relatively dry with low NDVI values resembling background properties. Tall, very green, tree species such as *Schotia brachypelata*, *Berchemia discolor* and *Spirostachys africana* displayed the highest NDVI ranges (with upper range limits reaching close to 0.9 and 1). The rest of the species had NDVI distributions that were severely overlapping making it difficult to separate the species further with the NDVI parameter. The PRI index was investigated next for patterns and possible thresholds.

The PRI parameter distribution (figure 3.7B) for all 23 tree species illustrated major overlapping ranges with no clear values serving as possible thresholds to separate the species. Majority of this species overlap was concentrated in the PRI range between ~-0.05 and ~0.02. *Combretum zeyheri* seemed to have the lowest range of PRI values with the lower limit reaching ~-0.09. The PRI parameter seemed to be the most ineffective of all the vegetation parameters and was thus not included in the decision tree design process. The CRI parameter was considered next.

The CRI histogram, in figure 3.7C, illustrated species specific value distributions that were more discrete and covered a narrower value interval than the previous vegetation indices. *Acacia nigrescens* had one of the lowest value ranges (0-~0.9), which were fairly distinctive, with a slight overlap with the lower range limits of other species. A significant degree of species overlap was evident in the CRI interval of ~0.1 to ~1.5. Certain tall tree species such as *Spirostachys africana* (~0.5-~2.7), *Schotia brachypelata* (~0.8-~4) and *Berchemia discolor* (~0.8-~3.5) had higher, more distinct CRI ranges with little overlap from other tree species. Despite the congested range of some species in the ~0.1-~1.5 interval, the CRI parameter proved useful especially in separating certain tall tree species while the separability of the smaller tree species still remained elusive. The final parameter, Red Edge NDVI (in figure 3.7D), will be discussed next.

Like the previous histograms, there was a certain degree of parameter overlap that existed for the various tree species, however, certain groups of species had distinct ranges which could be distinguished from others. Once again tall tree species such as *Diospyros mespiliformis*,

Schotia brachypelata and *Berchemia discolor* had distinctly high parameter ranges (~0.3-~0.7). The small tree species, and some bush encroaching species, which include *Combretum collinum, Pterocarpus rotundifolius, Terminalia sericea* and *Ziziphus mucronata* had lower distributions (~0.1-~0.43) which were significantly distinct despite little overlap with the tall tree species mentioned earlier. The two most common tree species, *Acacia nigrescens* (~0.6-~0.45) and *Sclerocarya birrea* (~0.1-~0.52), had strongly overlapping distributions across the ranges of most of the remaining tree species. Despite the level of overlap experienced in some of these histogram results, the spectral vegetation indices and minimum height values were subjected to a threshold formula (Pu et al, 2008) to create the species threshold matrix in table 3.3.

3.3.2 SPECIES THRESHOLD MATRIX

When the threshold formula (Pu et al, 2008) was applied, fixed lower and upper thresholds were created for each of the parameters for the different tree species. The trends evident in the matrix was similar to the results observed in the histograms but the parameter thresholds created a distribution or range that was more focused for each parameter and each species which facilitated decision tree node construction and associated branching. The species threshold matrix is display below:

Table 3.3: Species spectral and structural parameter threshold matrix

Tree Species	Vegetation Indices								Height	
	Lt NDVI	Ht NDVI	Lt PRI	Ht PRI	Lt CRI	Ht CRI	Lt RE	Ht RE	Lt Min	Ht Min
Acacia gerrardii	0.563	0.689	-0.016	0.011	0.649	1.023	0.319	0.434	1.602	3.359
Acacia nigrescens	0.372	0.545	-0.031	-0.014	0.269	0.584	0.184	0.301	5.096	8.618
Berchemia discolor	0.709	0.803	-0.022	-0.002	1.361	2.542	0.401	0.495	8.992	10.988
Combretum apiculatum	0.441	0.656	-0.056	-0.024	0.615	1.226	0.209	0.369	2.415	3.864
Combretum collinum	0.393	0.666	-0.053	-0.022	0.459	1.271	0.190	0.373	3.040	5.848
Combretum hereroense	0.517	0.715	-0.032	-0.002	0.622	1.299	0.278	0.442	2.190	4.015
Combretum imberbe	0.533	0.682	-0.028	-0.005	0.616	1.083	0.311	0.443	3.853	9.768
Combretum zeyheri	0.473	0.725	-0.062	-0.001	0.803	1.421	0.233	0.455	2.453	4.229
Dalbergia melanoxylon	0.301	0.452	-0.050	-0.022	0.224	0.552	0.143	0.246	2.062	2.670
Dichrostachys cinerea	0.497	0.659	-0.031	-0.005	0.550	1.042	0.255	0.378	1.734	3.556
Diospyros mespiliformis	0.631	0.734	-0.023	0.000	0.864	1.555	0.365	0.453	6.169	9.474
Euclea natalensis	0.519	0.804	-0.037	0.007	0.674	1.793	0.297	0.498	1.979	3.836
Gimnosporia senegalensis	0.556	0.718	-0.016	0.007	0.637	1.256	0.312	0.432	1.576	3.342
Lannea discolor	0.222	0.608	-0.056	-0.024	-0.106	1.279	0.082	0.354	2.871	4.936
Lonchocarpus capassa	0.492	0.711	-0.031	0.004	0.523	1.268	0.270	0.447	4.416	6.811
Peltophorum africanum	0.485	0.695	-0.038	-0.012	0.541	1.263	0.264	0.416	2.453	4.533
Pterocarpus rotundifolius	0.436	0.659	-0.038	-0.011	0.449	1.025	0.226	0.390	2.012	4.369
Schotia brachypelata	0.755	0.840	-0.008	0.014	1.490	2.917	0.466	0.552	6.931	8.723
Sclerocarya birrea	0.474	0.643	-0.033	-0.012	0.504	1.020	0.244	0.373	5.747	8.564
Spirostachys africana	0.653	0.805	-0.021	0.007	1.116	2.039	0.368	0.495	4.845	8.429
Strychnos sp.	0.541	0.763	-0.033	0.004	0.655	1.875	0.278	0.450	4.931	5.949
Terminalia sericea	0.455	0.630	-0.042	-0.006	0.504	0.994	0.217	0.343	1.836	3.982
Ziziphus mucronata	0.460	0.635	-0.027	-0.001	0.416	0.920	0.242	0.359	2.596	5.091

From the threshold matrix and the histogram results, it was evident that the minimum vegetation height, CRI and Red Edge NDVI parameter thresholds were the most useful in separating the tree species for the decision tree construction. For the minimum height parameter, it was clear that the 5 metre height threshold was suitable for dividing the species pool into the small and tall tree height categories. It was also evident that the species in the small tree height category (<5m) could be separated even further with the introduction of another minimum height threshold of approximately 3.8 metres. A NDVI threshold of 0.52, a Red Edge (RE) NDVI threshold of 0.43 and a CRI threshold of 0.95 were considered the most appropriate threshold values to break down the tall tree species into the various end nodes. Following the minimum height threshold of 3.8m for the small trees, a Red Edge NDVI threshold of 0.35 and a CRI threshold of 0.52 proved useful in splitting the small tree species into the various end nodes. From this series of minimum vegetation height and vegetation indices threshold values, a 9 end node decision tree architecture, with 7 intermediate nodes for branching, was constructed.

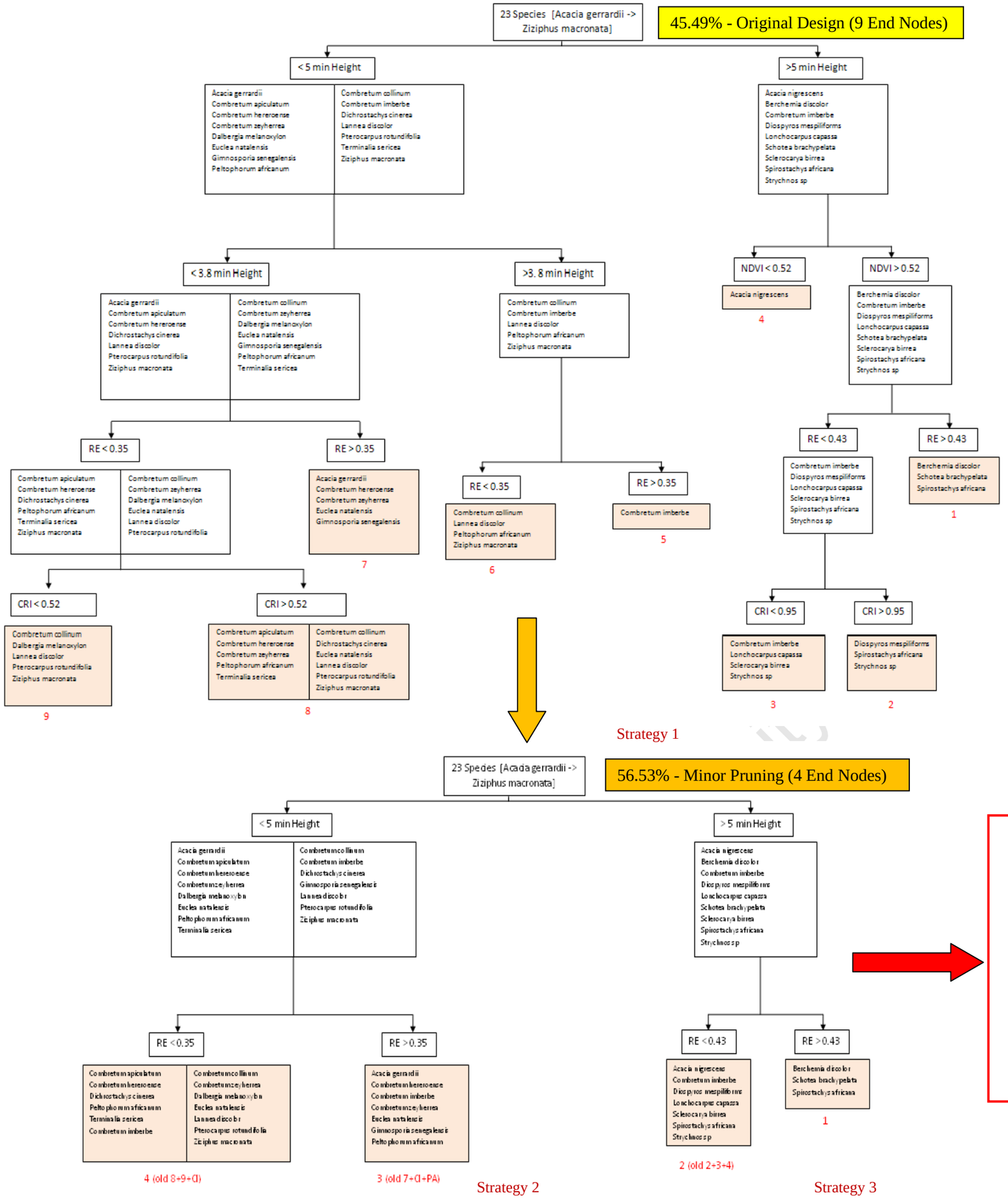
3.3.3 THE ORIGINAL DECISION TREE DESIGN CONFUSION MATRIX

The multiple endmember SAM classification and associated classification validation was conducted on the 9 end nodes (figure 3.4) which were incorporated to create a single confusion matrix (table 3.4) that displayed the users, producers and overall classification accuracy of the decision tree design. Individual end-node accuracies were considered for design improvement purposes but were not displayed in this results section as the overall classification accuracy, which tested the efficiency of the decision tree design, was the pivotal result worth deliberating.

Table 3.4: Original 9 End Node Decision Tree Confusion Matrix

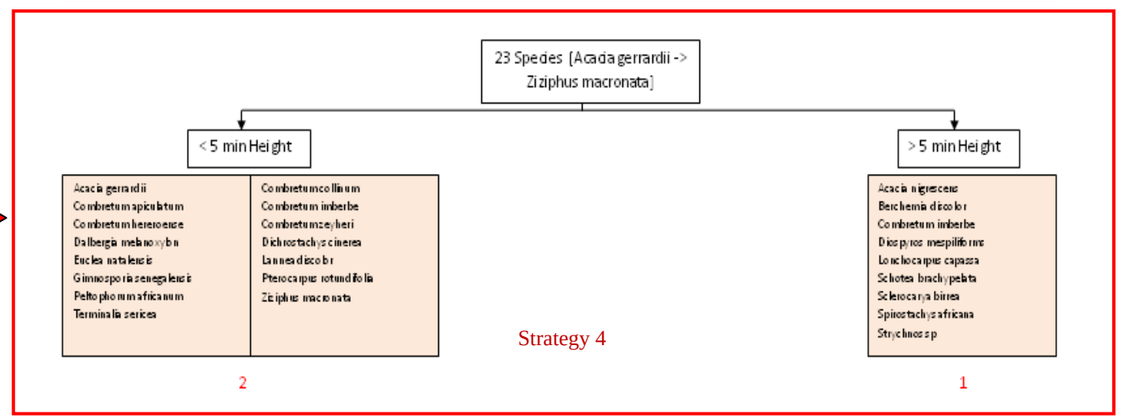
Confusion Matrix (9 End Node Design)	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	46			1	1	1	3			6		2					2	1	1			1	5	70	65.7
AN		159			13		8				1			3		5			43				18	250	63.6
BD			14								2							12						28	50
CA	16			41	8	4	4	3		9		1	1			4	1					1	7	100	41
CC	4	4		1	11	1	3	2			1	3				1	1	6	1				2	41	26.8
CH	2	2				16	2	1		1							2						4	30	53.3
CI	5	4			1		20	4	1	1	2	1	1					22	6					68	29.4
CZ	2				1		3	9		1								1						17	52.9
DalMel									6	2					1		2					1	2	14	42.9
DC	24	1		12	10	14	3	1	1	33		2	2				11	1	2			1	15	133	24.8
DM			1				1				11			2				12	4					31	35.5
EN	2					2	2			1		15										1	2	25	60
GS	1				1	2	1	1		2		3	5			1							1	18	27.8
LC	3	4					7		1					7				7	1				1	31	22.6
LD							2					1			7			2						12	58.3
PA	6	2		1	1	1	4		2			1	2			3		1					2	26	11.5
PR	4	3		2	4	6		2	1	1		2	1				17		1					44	38.6
SA			1								3							72	3	2				81	88.9
SB	1	80	1	1			7				36			5			1	33	106	2		1		274	38.7
SchotB																		4		11				15	73.3
Stry																		1		3	5			9	55.6
TS	4			2	4	3				3					1	1	2		1			18	14	53	34
ZM	5	3		1	2	2	3	2	2	1		1	1	1		1	1	3	1			1	10	41	24.4
Grand Total	125	262	17	62	57	52	73	25	14	61	56	32	13	18	9	16	40	178	170	18	5	25	83	1411	
Producer's Accuracy	36.8	60.69	82.35	66.1	19.3	30.77	27	36	42.857	54.1	19.6	47	38	39	78	19	43	40	62	61.111	100	72	12		
Overall Accuracy	45.49																								

The classification, based on the original 9 end node decision tree design, yielded fairly poor results with the overall classification accuracy being 45.5% (less than 50%) for the 23 tree species. This is the result of a lot of classification confusion between the particular species. Classification confusion between species resulted as the two or more confused species share particularly similar spectral and possible structural traits which made it difficult for the SAM classifier to differentiate between them. According to the producer's accuracy (indicative of the error of omission), at an individual species level, some species yielded accuracies greater than 60%, such as *Combretum apiculatum* (66%), while others yielded accuracies greater than 70% such as *Berchemia discolor* (82.4%), *Lannea discolor* (78%) and *Strychnos species* (100%). The results for *Strychnos species* were not considered a reliable result due to an under-representative validation sample (only 5 validation pixels). The remaining species obtained insignificant results with 10 out of the total 23 species yielding classification accuracies lower than 40% and one, namely, *Ziziphus mucronata* yielding the lowest accuracy of 12%. The performance of *Ziziphus mucronata* was attributed to the spectral confusion, mostly, with three other species: *Acacia nigrescens*, *Dichrostachys cinerea* and *Terminalia sericea*. Some of the bush encroaching species were effectively classified such as *Terminalia sericea* (72%) and *Combretum apiculatum* (66%) while others yielded poorer results such as *Acacia gerrardii* (36.8%). The poor result observed for *Acacia gerrardii* was due to the classifier spectrally confusing the species mostly with *Dichrostachys cinerea* (another bush encroacher species) in which 24 *Acacia gerrardii* pixels were classified as this species. The two most prevalent and sampled savanna tree species in the study region, *Acacia nigrescens* and *Sclerocarya birrea*, obtained a result of 60.7% and 62% respectively which had to be improved significantly. Majority of the classification confusion of *Acacia nigrescens* was with *Sclerocarya birrea* and vice versa. The user's accuracy (a measure of the error of commission) also yielded poor results with the majority of the species yielding low accuracies due to the incorrect classification of additional pixels to the different species classes. The highest user's accuracy was obtained by *Spirostachys africanus* with 88.9%, which was followed by *Schotia brachypelata* with 73.3%. All bush encroaching species, with exception to *Acacia gerrardii* (65.7%), yielded very low user's accuracies which indicated a great deal of spectral confusion experienced by the SAM classifier when implementing this particular decision tree design. This original 9 end node decision tree had to undergo various redesign steps (according to four main strategies) to improve the accuracy results which is summarised in the decision tree evolution schematic (figure 3.8) below.



3.3.4 EVOLUTION OF THE ORIGINAL DECISION TREE DESIGN (FIGURE 3.8)

- By conducting a minor pruning strategy (4 end nodes), the overall accuracy improved from **45.49%** to **56.53%. [<u>Aside:</u> Refer to Appendices 3B to 3E for the strategies and the baseline result confusion matrices]**
- In the minor pruning decision tree design, end node 1 remained the same while end node 2 was reconstructed from the combination of the old end nodes 2, 3 and 4. The new end node 3 comprised of the old end node 7 plus the addition of *Combretum imberbe* and *Peltophorum africanum*. The final end node 4 was made up of old end nodes 8 and 9 plus *Combretum imberbe*. The old end nodes 5 and 6 were excluded entirely.
- Individual species classification accuracies all showed signs of massive improvements especially for some of the problematic, spectrally confused species (e.g. *Diospyros mespiliformis* with a jump from 19.6% to 73.7%). Some species did, however, illustrated a decline in accuracy in comparison to the previous design (*Dichrostachys cinerea* with a fall from 54.1% to 43%).
- This minor pruning strategy and design was not considered the most optimum as it still fell short of the overall accuracy obtained by the traditional multiple endmember SAM baseline result of **56.67%**.
- Among all the implemented strategies, the utilisation of selected training endmembers, instead of the entire training dataset (~1486 spectra), for the SAM classification process yielded the worst overall accuracy of **49.75%**.
- In contrast, the utilisation of selected bands yielded a more positive result of **56.64%** which was marginally behind the baseline result. Due to the positive result of this strategy, these selected bands were also implemented in the extreme pruning strategy next.
- The **extreme pruning strategy (2 end nodes)** yielded the highest overall classification accuracy of **62.18%** and was considered the **final decision tree design** for this study. This decision tree design did not utilise any vegetation indices parameter thresholds and simply branched the species pool according to a single minimum height threshold (5m).
- Species (such as *Combretum sp.*) were aggregated into a single family class for classification to boost accuracy results. The improvement was insignificant and excluded from the results.



62.18% - Extreme Pruning (2 End Nodes) + Selected Bands

VERSUS Baseline Result

56.67% - Traditional Multiple Endmember SAM + Selected Bands

3.3.5 FINAL DECISION TREE DESIGN CONFUSION MATRIX

Table 3.5: Final 2 End Node Decision Tree Confusion Matrix

Confusion Matrix (2 End Node Design)	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	51	1		3	1	2	3			6						3	2		2					74	68.9
AN	3	171		1	3	3	9		6	6				4	1		2		34			1	3	247	69.2
BD			20								1			1				4	1	1				28	71.4
CA	7			60	4	4	2	1		10			1				1					5	5	100	60
CC	1	3	1	4	11	1				2			2	2		2	3	3	4					39	28.2
CH	2	2				20	2			2							1						1	30	66.7
CI	5	5				2	41			5			1	1				4	4					68	60.3
CZ	1			1			2	11		1												1		17	64.7
DalMel				1					8	1						1						2	1	14	57.1
DC	14	3	1	14	3	4	10			69		4				2	10					1	5	140	49.3
DM							1				26							2	2					31	83.9
EN	1			2						1		18				3								25	72
GS	2			1		2	2			3			6			2	1							19	31.6
LC	3	7					3			1	2	1		9	1		1	2				1		31	29
LD								1	1				1		7			2						12	58.3
PA	3					1	7			3	1		1			6	1		1			1	1	26	23.1
PR	1	2		3	3	2	2		1	5		2				2	20		2			1	1	47	42.6
SA		1	7				1				3							63	3	2	1			81	77.8
SB	1	38		1			14				2			5		1		7	204			1		274	74.5
SchotB																		2		13				15	86.7
Stry																		1		2	6			9	66.7
TS					1	4	2			8		1					1		1			34	2	54	63
ZM	3	2		2		2	5	1	3	5							2	2	2			3	12	44	27.3
Grand Total	98	235	29	93	26	47	106	14	19	128	35	26	12	22	9	22	45	92	260	18	7	51	31	1425	
Producer's Accuracy	52	72.77	69	64.5	42.3	42.6	38.7	78.6	42.10526	53.91	74.29	69.2	50	40.9	77.8	27.3	44.4	68.5	78.5	72.22222	85.71	66.7	38.7		
Overall Accuracy	62.18																								

The producer's accuracy results, at an individual species level, obtained from the extreme pruning decision tree design were significantly higher than the previous 9 end node design with a significant reduction in the classification confusion that were evident between certain species. The *Strychnos species* was suitably classified with an accuracy of 85.7%. The two most common tree species, *Acacia nigrescens* and *Sclerocarya birrea*, were classified at high accuracies (72.8% and 78.5% respectively). The spectral confusion between these two species, which was a noticeable problem before, was still present but was significantly reduced. The classification accuracies for the bush encroaching species were more or less at the same levels as before with a few species showing minor increases and decreases. The most noticeable was the decline in accuracy for *Terminalia sericea* which dropped from 72% to 66.7%. On the other hand the poorest performing bush encroaching species, *Acacia gerrardii*, showed a significant improvement in classification accuracy from 36.8% to 52%. Although not adequate by any means, in comparison to some species, the results of all the bush encroaching species were relatively positive. There was no significant improvement for *Pterocarpus rotundifolius* and *Dalbergia melanoxylon* with classification accuracies still less than 50% (44% and 42.1% respectively). Other tree species which showed vast improvements in classification accuracies, under the final decision tree design, included *Combretum zeyheri* (36% to 78.6%), *Diospyros mespiliformis* (19.6% to 74.3%), *Euclea natalensis* (47% to 69.2%), *Spirostachys africana* (40% to 68.5%) and *Schotia brachypelata* (61.1% to 72.2%). Under the final decision tree design, 15 out of a total of 23 species yielded classification accuracies greater than 50% with 12 out the 15 achieving accuracies of 65% and greater. Although improvements in accuracy were observed for most species, some of the poor performing species were *Combretum imberbe* (38.7%) and *Ziziphus mucronata* (38.7%) with *Peltophorum africanum* yielding the lowest classification accuracy (27.3%). *Peltophorum africanum* displayed strong spectral confusion with *Acacia gerrardii* and *Euclea natalensis* while *Combretum imberbe* was mostly confused with *Sclerocarya birrea* (14 misclassified pixels) during the classification process.

In terms of the user's accuracy, the accuracies, for the different species, showed equally large improvements as the producer's accuracies. The species with the highest probability of being correctly represented at the ground level was *Schotia brachypelata* (86.7%) with *Diospyros mespiliforms* (83.9%) following closely behind. As with the producer's accuracy results, *Peltophorum africanum* also achieved the lowest user's accuracy of 23.1%.

With bush encroaching species being the primary focus of this thesis, the bush encroaching species' distributions were extracted from the classified end node images and displayed as individual species maps in the final section up next.

3.3.6 BUSH ENCROACHING SPECIES MAPS OF THE L456 REGION

Based on the classification results, from the confusion matrix (table 3.5), only 4 out of the 6 bush encroaching species maps, which illustrated classification accuracies greater than 50%, were displayed. These species included *Acacia gerrardii*, *Combretum apiculatum*, *Dichrostachys cinerea* and *Terminalia sericea*. These maps were displayed in conjunction with the original unclassified hyperspectral image (figure 3.9) which illustrated a few geographical areas of interest such as nearby informal settlements, fence lines and main rivers etc.

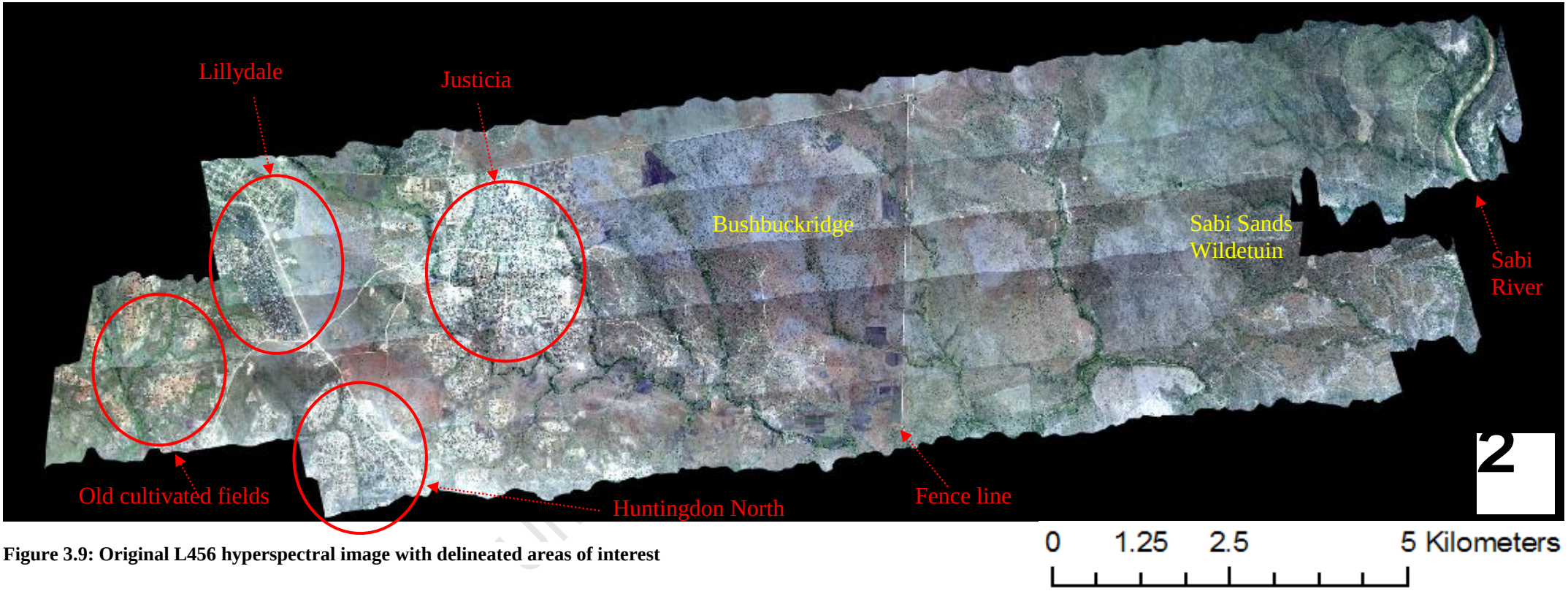
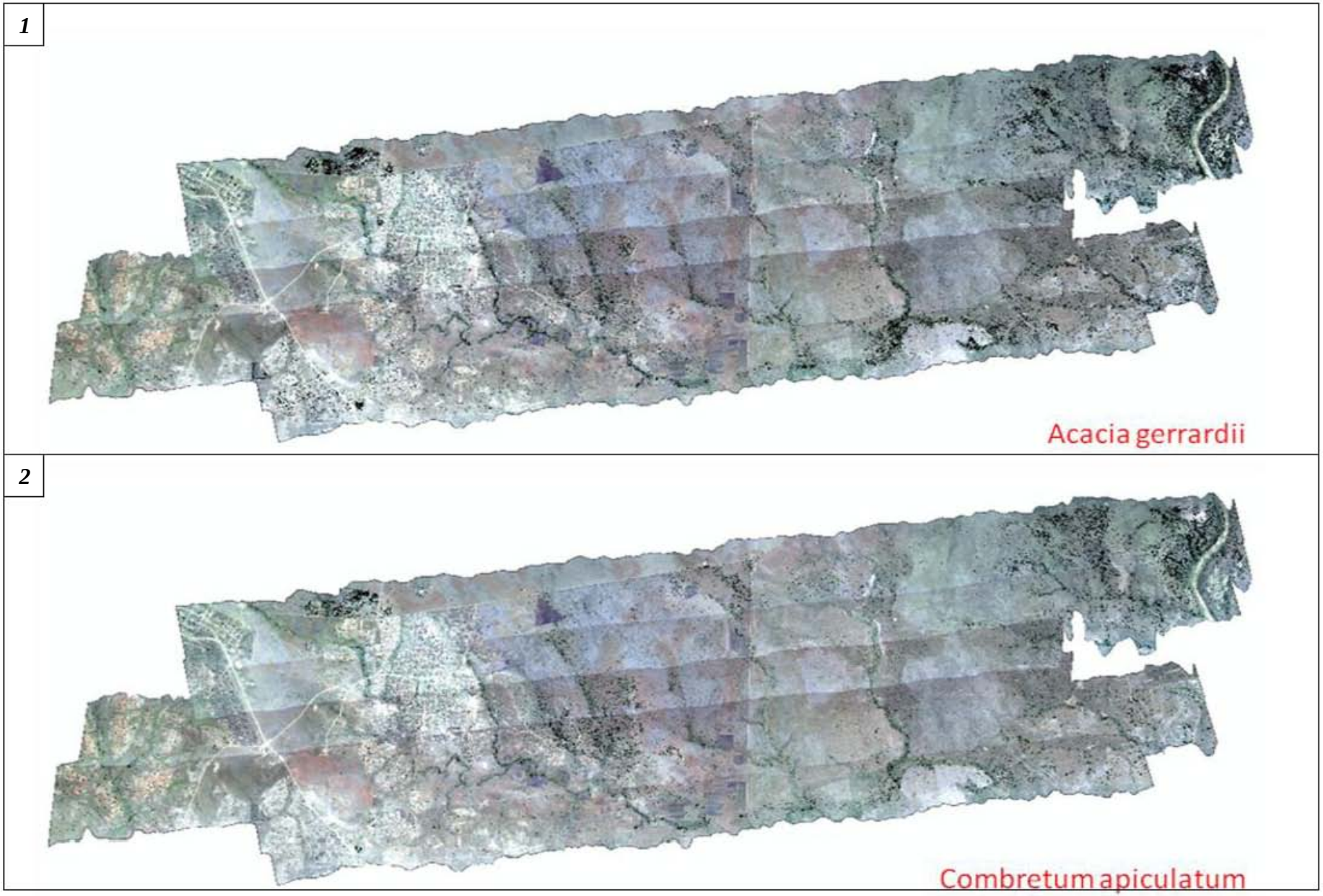
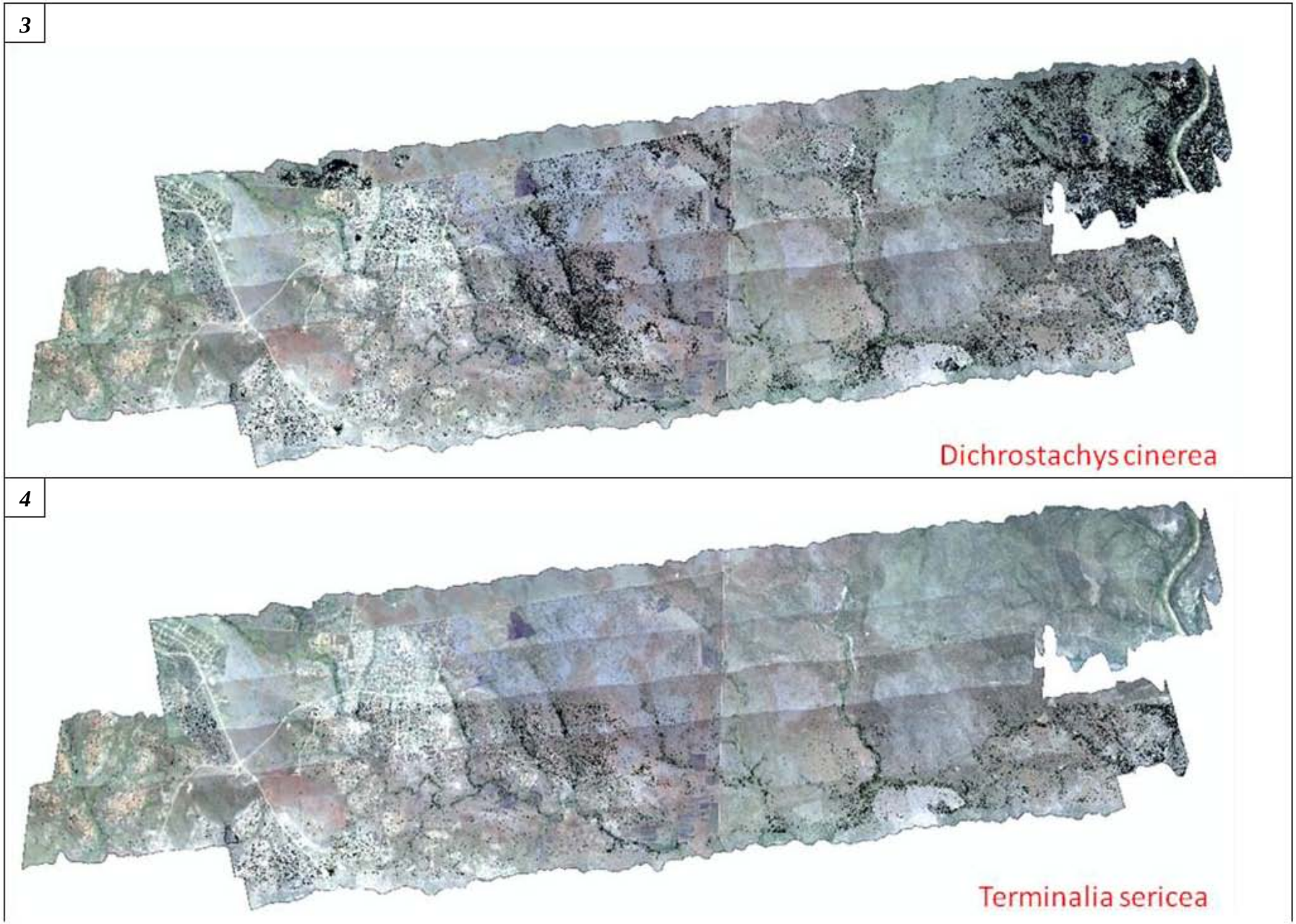


Figure 3.9: Original L456 hyperspectral image with delineated areas of interest



Figures 3.10 & 3.11: Species distribution maps of bush encroachers – (1) *Acacia gerrardii* & (2) *Combretum apiculatum*



Figures 3.12 & 3.13: Species distribution maps of bush encroachers – (3) *Dichrostachys cinerea* & (4) *Terminalia sericea*

The bush encroaching species *Acacia gerrardii (figure 3.10)*, *Combretum apiculatum (figure 3.11)* and *Dichrostachys cinerea (figure 3.12)* all followed a spatial pattern which was fairly similar but differed in terms of density. These species illustrated a distinct dense distribution in the Northern Sabi Sands region with distributions emanating from the riparian zone along the Bushbuckridge landscape. The series of troughs and crests just to the east of Justicea and the vegetated area to north showed dense clustering of these three bush encroaching species. The pattern of *Combretum apiculatum* differed from the other two similarly distributed species with very limited presence in the southern extents of Sabi Sands and the communal rangelands near Lillydale, Northern Huntingdon and the old cultivated fields. It was also evident that *Combretum apiculatum* illustrated the least intense spatial distribution with *Dichrostachys cinerea* displaying the most intense patterning across the study region. *Dichrostachys cinerea* was well distributed throughout the study region with the encroaching species occurring in most areas in and around the rangelands and Sabi Sands. In contrast to the other species, *Dichrostachys cinerea* occurred extensively in the riparian zone of the Sabi River. The fence line visibly restricted the distribution of this species from reaching the northern extents towards the privately owned game reserves.

Terminalia sericea (figure 3.13), on the other hand, illustrated a spatial pattern that was very distinct relative to the rest of the bush encroaching species. This species' distribution was strictly limited to the southern extents of the study region with no major distribution evident in Northern Sabi Sands and northern extents of the rangelands. Interestingly this species was found not to occur in areas to the north of Justicea and in the northern reaches of Lillydale where the other bush encroaching species were quite prominent. Majority of this species distribution was concentrated in the area east of Justicea and in the Southern Sabi Sands region and along the southern riparian tributaries. *Dichrostachys cinerea* and *Terminalia sericea* were the only two bush encroaching species which occurred noticeably within the actual informal settlement of Justicea. At the more general scale, the only area in the North Sabi Sands region which was devoid of any of the bush encroaching species was a large crested feature just to the west of the Sabi River. In the communal rangelands, the bush encroaching species were the least prevalent in areas such as the old cultivated fields, Northern Huntingdon and Lillydale and on major crests along the Bushbuckridge landscape.

3.4 DISCUSSION

The real strength of the decision tree methodology, over other methods, was the incorporation of multiple useful parameters or datasets which could collectively aid the pixel based classification process. The histograms and threshold matrix results gave the user a rough gauge of which parameters proved useful for incorporation into the decision tree design and the possible thresholds to separate the various species' value ranges. They also gave an early indication of which of the different species would be easily discriminated during the classification process and which species would prove to be a challenge. This type of preliminary analysis does not only allow the assimilation of only the most useful available data for the decision tree construction process but it also offers an indication of where to begin first.

From the multiple variable histograms and threshold matrix results, it was evident that the minimum tree height parameter played a significant role in separating the species pool into distinct groups than the individual vegetation indices parameters alone. The tall tree species, especially *Acacia nigrescens* (in most cases), *Diospyros mespiliformis*, *Schotia brachypelata* and *Berchemia discolour*, were easily the most separable in the different parameter histograms. The smaller tree species (e.g. *Dalbergia melanoxylon*, *Combretum collinum* and *Peltophorum africanum*, just to name a few) and problematic species (*Combretum imberbe*) were the most difficult to separate due to the high degree of spectral and structural parameter overlap. *Combretum imberbe* was identified as a 'problematic' species during the early stages of analysis because the species had a minimum height range that covered most of the ranges expressed by most species (both small and tall tree ranges) and the spectral properties, according to the vegetation indices parameters, were not distinctly discernable for possible differentiation. This lack of a certain species or species' spectral and structural separability is the direct cause of most of the classification confusion that would be experienced by the SAM classifier. Additionally, decision tree design flaws, such as the addition of unnecessary nodes, could also contribute to this classification confusion. It was expected that the strength of the multiple endmember SAM classifier would be sufficient to spectrally separate these small tree and problematic species if the decision tree design was optimal. The effectiveness of this classifier and decision tree design, assessed by means of confusion matrices, will be critiqued at a later stage.

For the original 9 end node decision tree design, it was recommended that the minimum height parameter be utilised first to branch the species pool, at a rough threshold of 5m, into two root nodes. The 5 metre height threshold was chosen as it was a lenient boundary which allowed for the creation of two distinct species height groups (short and tall tree groups) without cutting off much of the natural height variability of the different individual tree species. This 'safe' height threshold of 5m was supported by the trend in the histogram figure 3.6 and the species threshold matrix in table 3.3. This simple branching also significantly helped reduce the species parameter overlap evident in the vegetation indices histogram results. From the histogram and threshold matrix results, it was also clear that the Red Edge NDVI and CRI parameters were the most useful spectral parameters for incorporation into the decision tree architecture to separate the species even further. The effectiveness of these two vegetation indices could have been the result of utilising the significant bands in primary and secondary plant chemical spectral regions, identified in the previous chapter. The PRI vegetation index was the most ineffective in this regard with no evidence of discernable thresholds for separating the different species. The PRI index was thus excluded from the decision tree construction process. The trends observed in the Red Edge NDVI were more amenable to the establishment of possible thresholds than the ordinary NDVI parameter and thus the NDVI index was also excluded. When constructing the original decision tree design, the overarching approach was to separate and branch off the species pool, as much as possible, into the different end nodes that contained the minimum number of possible species. This approach would help reduce the strain experienced by the multiple endmember SAM classifier during the classification process.

The minimum height, CRI and Red Edge NDVI parameter thresholds obtained from the threshold matrix were chosen to maximize the separability between the different species ranges and at the same time not distort the species parameter ranges too excessively when branching them off into the particular intermediate nodes or end nodes. In terms of the thresholding process, the use of the a single standard deviation interval in the Pu et al (2008) threshold formulae created a threshold range that captured the majority of the species' parameter variability. In this study, preliminary trials illustrated that the implementation of a doubled standard deviation interval would create species parameter threshold ranges too large and overlapping to adequately separate during node construction. The implementation of a half standard deviation interval in the formulae would have created more discrete parameter threshold ranges which would make separation and node construction far easier. These

thresholds, however, would not be truly representative of the variability contained within the parameter data, present in the histogram results, for the different species. A species parameter range that was not fully covered by the threshold value at a particular node would results in the introduction of classification error for the species which would compromise the overall accuracy of the decision tree design. Unfortunately the use of fixed threshold values for parameters at particular nodes in the decision tree design could not fully encompass the parameter ranges of all the species in the species pool and thus a certain degree of classification error was unavoidable.

After running the SAM classifier script, the original 9 end node decision tree design (figure 3.4), though seemingly very comprehensive in accounting for the specific species parameter ranges, yielded the lowest overall classification accuracy in this study (45.5%) with high classification confusion between particular species. In an attempt to generate end nodes that possessed the minimum possible number of species to aid the SAM classification, the error accumulated during the branching process along the total of 16 nodes (7 intermediate nodes + 9 end nodes) was suspected to be relatively high. Another possible reason for the poor performance was that the chosen parameter thresholds may have been too strictly implemented during the decision tree construction. Certain species, which displayed overlapping parameter ranges, may have been incorrectly lumped into one particular node instead of co-existing across multiple nodes within the same tier. The rough accuracy statistics for each end node and the overall confusion matrix also demonstrated the lack of effectiveness of particular end nodes at the species level which prompted a decision tree redesign. Also from the confusion matrix and from the histogram results, various problem species were identified (e.g. *Combretum imberbe*, *Ziziphus mucronata* and *Combretum collinum* to name a few) which possessed high degrees of classification confusion and very low accuracies due to certain design flaws that needed remedying. Four main strategies were implemented to improve the decision tree design and reduce the overall classification error and the classification confusion between species (figure 3.8). Regardless of the strategies that were considered, it was clear that a simpler design with less branching and total end nodes was the way forward for this study.

The first strategy was the minor pruning strategy which yielded improved results (from 45.5% to 56.5%) when reducing the decision tree end nodes from 9 to 4. The minor pruning strategy was introduced to streamline the design by removing the various decision tree design

flaws that were evident from the original design confusion matrix result (figure 3.4). Such design flaws in the original design included the creation of end nodes 2 and 3 which resulted in high spectral confusion between *Diospyros mespiliformis* and *Sclerocarya birrea* with *Diospyros mespiliformis* obtaining very low classification accuracy. End node 4, which isolated *Acacia nigrescens*, caused significant spectral confusion with *Sclerocarya birrea* which resulted in low accuracies for both species. By combining the species in both nodes 2 and 3, the classification accuracies, especially for *Diospyros mespilifomis*, improved greatly. Another similar example was the creation of end nodes 5 and 6 to isolate *Combretum imberbe* which yielded particularly poor results. Reverting back to the intermediate node, in the previous tier, also improved the results fairly significantly for the species concerned. The second minimum height threshold (3.8m) parameter and its subsequent nodes was also scrapped from the design as it hinged on the success of establishing a separate end node for *Combretum imberbe*. Since the other species had overlapping ranges across the 3.8m minimum height threshold nodes, the change in design had no adverse effects on the species in that tier. End nodes 8 and 9 were also excluded for the sake of a simpler design. As a final measure to reduce error, the parameter thresholds at each node were relaxed in order to accommodate overlapping species, such as *Combretum imberbe* and *Peltophorum africanum*, across multiple end nodes in the decision tree. The new decision tree design (figure 3.8) was limited to a total of 4 end nodes which were populated with a greater number of species. This reduction in the number of end nodes and simplification of the decision tree design had placed more pressure and reliance on the SAM classifier, in terms of computing requirements and time, to distinguish the species at the individual species level. The results, however, proved positive (Appendix 3C) and even though the new design did not obtain an overall classification accuracy greater than the standard multiple endmember SAM result (figure 3.8 & Appendix 3F), the 4 end node design was implemented as the base onto which the remaining strategies were applied in an attempt to achieve the highest overall accuracy.

Instead of using the entire training dataset for the multiple endmember SAM classifier, a strategy involving the selection of representative training endmembers, via cluster analysis, was implemented. Amongst all the implemented strategies, this strategy yielded the lowest overall accuracy result (figure 3.8 & Appendix 3D). The poor performance was attributed to the spectral endmember selection process in which the endmembers were extracted from various species specific dendrograms that were constructed from the training reflectance values. In the dendrograms, spectra that were representative of the distinct groupings or

clusters within the particular species training dataset were chosen based on the user's interpretation. The incorrect spectra or lack of a sufficient number of representative spectra may have been utilised as endmembers for particular species. This might have been the case with particular species that had enormous training datasets (e.g. *Sclerocarya birrea* and *Acacia nigrescens*) in which the clustering were too congested for adequate interpretation. Another possible reason for the failure of this strategy was the utilisation of the inappropriate linkage rule in the cluster analysis approach. A single linkage cluster analysis was utilised to create the dendrograms which illustrated different clustering patterns in comparison to other types such as complete linkage, pair group averages and weighed variants. One of the other linkage rules may have produced better results for easier and more appropriate endmember extraction. Ascertaining the appropriate linkage rule and the appropriate number of sufficiently representative endmembers for each species was considered a whole task on its own and thus the strategy was discontinued.

The final strategy that was applied to the 4 end node design was the selection of significant bands to optimize the classification capabilities of the SAM classifier. The bands were selected based on the results from the previous chapter and the bands utilized in the Cho et al (2010). The results of this strategy surpassed the performance of the previous strategies with an overall accuracy marginally lower than the baseline result. This strategy was also considered advantageous for future design changes as it played an enormous role in reducing image data dimensionality (from the original 72 bands to 31 significant bands) which sped up processing time and reduced computational demand. This result confirmed one of the outcomes of the Cho et al (2010) study, which concluded that the use of selected bands were performed better than using the entire 72 band spectrum as it contained the bulk of the potential species discriminating power. It was also thought, at this stage, to combine or bin species classes into a broader family class in an attempt to improve accuracies. This was only applicable to the Combretum species in the species pool which obtained miniscule improvements in the overall accuracy result. Since *Combretum apiculatum* was identified as an important bush encroaching species for analysis, this aggregation or class binning strategy was not warranted. With the selected band strategy, the 4 end node design reached the limit of its classification accuracy potential (56.6% - Appendix 3E). To get the overall classification accuracy into the levels of 60% and above, further revision of the decision tree design was necessary.

A two end node decision tree design was proposed based on the extreme pruning strategy in which the species pool was split up according to the 5m minimum height threshold only without any influence from the spectral vegetation indices parameters. Because of the positive performance and benefits of the selected band strategy, the selected significant bands were also implemented in the classification process of this strategy. It was determined that the most simplistic of decision tree designs (figure 3.8) actually yielded the highest overall classification accuracy, of 62%, in this study. This two end node design was thus considered as the final decision tree design for the mapping procedure. This design confirmed the trends observed in the histogram section of the results which emphasised the importance of the minimum height parameter as the main discriminator between the different tree species. The lack of influence of the vegetation indices parameters on the final decision tree design could indicate that these indices were not the most effective indices available to fully exploit the discriminating potential present in the significant bands. If the shortwave infrared (SWIR) region was available when the CAO sensor was flown, indices which made use of this region would have played a more dominant role in aiding species discrimination, decision tree design and mapping than the indices used in this study. The importance of the SWIR region was supported by the mapping process in the Asner et al (2008 – b) and Asner & Heidebrecht (2002) studies.

When examining the associated confusion matrix result (table 3.5), it was evident that all 23 species experienced significant improvements (some more than others) with also a significant reduction in the spectral confusion experienced between certain species. Some tree species which were earlier considered 'problematic', such as *Diospyros mespiliformis* and *Combretum zeyheri*, improved drastically to high levels of classification accuracy. Even after the implementation of the numerous strategies and changes in decision tree design to find the most optimum accuracy results, certain tree species still proved too challenging for classification at accurate levels. These species included mainly *Combretum imberbe*, *Peltophorum africanum* and *Ziziphus macronata* with producer's classification accuracies less than 40%. Although the user's accuracy for *Combretum imberbe* was greater than 60%, the other two species yielded poor user's accuracies (<30%) which indicated that these species were poorly represented and classified at the ground level. The reasons for the poor performance could be species specific but it was evident from the analyses that these species exhibited spectral characteristics that were too similar to be sufficiently distinguished from other species in the dataset. For *Peltophorum africanum* and *Ziziphus macronata* tree

species, the poor results could be related to user error, during the data collection and pre-processing stages of analysis, in which spectrally impure or 'mixed' spectra were collected and implemented as endmembers. Mixed spectra are especially common in tree species which have crowns that were relatively small and/or sparse with patches of the bare soil or grassy under-storey being captured during the spectra collecting process. An insufficient number of large enough canopies for pure spectra collection could also contribute to the species' low accuracies. The poor results of *Combretum imberbe* was mainly due to the fact that it possessed high structural variability with a minimum height range which overlapped significantly with most of the species in the species pool. This lack of structural separability, together with the underperforming spectral vegetation indices parameters, resulted in an overall lack of species discriminating ability for *Combretum imberbe* during the classification. The bush encroaching tree species maps were discussed next.

As a welcomed departure from the conclusions of the previous chapter regarding bush encroaching species, which stated the imminent lack of inter- and intra-species spectral differentiation, this mapping approach made the mapping of these species possible. The maps of *Acacia gerrardii*, *Combretum apiculatum*, *Dichrostachys cinerea* and *Terminalia sericea* were displayed as they yielded accuracies greater than 50%. The *Dalbergia melanoxylon* and *Pterocarpus rotundifolius* maps, on the other hand, were excluded as their distributions were not credibly represented. From the mapped results of the individual species, it was clear that the majority of the distribution was concentrated in and around the riparian zones, especially across the communal rangelands and Sabi River. It was postulated that this trend was the result of aggressive seed propagation in which the encroaching species may have originated from outside areas but was carried into the area by and established along the water courses. Amongst the bush encroaching species, *Dichrostachys cinerea* was evidently the most prolific encroacher with a dense distribution which was relatively unrestricted by differences in land use or the localised rainfall gradient in L456. Besides its wide scale persistence, this species was considered the most threatening of the bush encroaching species for a number of reasons. This spiny and shrubby tree is not a preferable source of fuelwood for the local populace and it directly impedes the movement and grazing capacity of the herders' livestock. This species also persist in areas of high human related disturbance as they are found invading the neighbouring informal settlements. In the Sabi Sands region, this species severely congests the northern portions of the reserve which could significantly reduce the populations of grazers and also obscure game viewing opportunities

in certain areas thus negatively impacting eco-tourism. The fence line seemed to prove effective in containing this dense distribution mostly within the communal ranges and preventing the species (and other encroachers) from reaching the crests and mid-slopes of the western Sabi Sands region. *Combretum apiculatum*, on the other hand, was the least prolific bush encroaching species in the study region. This species should not be considered to be a major threat for management, at the moment, due to its modest distribution across the landscape. Concentrated patches of this species in the northern Sabi Sands region, however, would require careful monitoring in case this species becomes as rampant as *Dichrostachys cinerea*. The same should be considered for *Acacia gerrardii* which displayed a spatial distribution pattern that was fairly similar to *Dichrostachys cinerea* but was much less intense. *Terminalia sericea* was the only encroaching species which displayed a distinct distribution pattern which was restricted to the southern extents of L456. This species appeared to occur especially near and around informal settlements and other areas of high disturbance. *Terminalia sericea* is a valuable source of timber and is a favourable fuel wood species in Bushbuckridge which is actively harvested by the local populaces for energy purposes (Dr B. Erasmus, personal communication – April 2009; Schmidt et al 2007). When cut these species coppice aggressively and are actively propagated in these high disturbance areas which is usually near and around informal settlements. The southern distribution of *Terminalia sericea* may be attributed to the fact that it is a seep-line species and tend to avoid the water scarce areas in the northern reaches (Dr E. Witkowski, personal communication – September 2010). The species also has an affinity for deep sandy soils and are not found on the more clayey soils. Next, possible strategies, based on these species distribution maps, were put forward for discussion which could significantly help both park and communal rangeland management.

Since being a valuable target species by the local population, the *Terminalia sericea* distribution map could be very helpful for communal rangeland management as encroached patches of this species could be identified and earmarked for sustainable harvesting on a rotational basis to prevent overutilization and thus possible bush encroachment via coppicing. This strategy could be implemented in and around the informal settlements of Justicea, Lillydale and Huntingdon North to prevent the local population from walking great distances to obtain their fire wood. The densely encroached area (containing *Acacia gerrardii*, *Combretum apiculatum* and *Dichrostachys cinerea*) to the north of Justicea could be targeted for clearing to free up land which could be utilised for grazing by the livestock herders.

Mechanical and chemical based clearing methods could be instilled to remove the encroachers but success is always variable as no particular methods have been proven to succeed 100% of the time (as is the case in Joubert, 2007). Similar efforts need to be conducted in the northern Sabi Sands region if eco-tourism or grazer populations were found to be suffering due to the congestive persistence of *Dichrostachys cinerea*. Bush encroaching species, however, are viewed both in the positive and negative light and therefore their associated management strategies are heavily dependent on the land management type (communal versus protected) and by weighing the socio-economic versus ecological needs in the study region. These species distribution maps ultimately provide management with the appropriate tools necessary for decision making at the regional scale.

3.5 CONCLUSIONS

It was concluded that this novel decision tree multiple endmember SAM classification approach was more effective in discriminating and mapping savanna tree species than a traditional straight forward multiple endmember SAM approach. During the decision tree design process, it was clear that the minimum tree height parameter played the greatest role in separating the species pool, prior to species level classification, than any of the spectral parameters (NDVI, Red Edge NDVI, PRI and CRI) considered for the study. When the different strategies and associated decision tree designs were reviewed, it was conclusive that the optimum design was one which utilised only two end nodes constructed from the 5 metre minimum tree height threshold. This design together with spectral subsetting, by making use of selected significant bands, yielded the most attainable results for this study. In this case, it was the simplest design possible (2 end node) which prevailed over the more complex. Even though many of the species in the species pool illustrated classification accuracies of 60% and greater, certain species underperformed and required significant improvements in their classification accuracies before their distributions could be considered plausible (e.g. *Combretum imberbe* and *Peltophorum africanum*). From the bush encroaching species maps, the encroaching species were found to be concentrated in particular areas across the L456 study region. These 'encroached' areas include the entire Northern Sabi Sands region, the riparian zones, the mid-slopes and troughs landscape features to the east of Justicea and the area to the north of Justicea. *Dichrostachys cinerea* was the most dominant encroaching species with the most prolific distribution while *Combretum apiculatum* was the least

abundant. The dense distribution of *Dichrostachys cinerea* was alarming and could be of major concern to management and should thus be earmarked for future monitoring. *Terminalia sericea* had the most distinctive distribution pattern, in comparison to the other encroacher species, which was constrained to the southern extents of the study region. At the end of the study, it was evident that this newly proposed method yielded classification accuracies and tree species maps of a particularly reasonable calibre but certain improvements and suggestions were recommended to create a foundation for future work. It is also important to note that this novel method can be equally applicable in the mapping of economically viable savanna tree species, tall trees species, alien invasive vegetation and other species of interest found across the savanna landscape. Although this newly proposed method was developed only from the spectral and structural information extracted from the L456 imagery, it is repeatable in any type of savanna environment only if the appropriate season-related spectral endmembers (i.e. endmembers encompassing the spectral properties and variability of the tree species at the time of image acquisition) and tree height information were utilised for the decision tree construction and image classification. It was also conclusive that this new method could make major contributions to ecosystem management in general as an important tool for managing and monitoring different land use areas at the finest and closest of scales which would lead to more sound and species specific decision-making.

CHAPTER 4: THESIS SYNTHESIS AND RECOMMENDATIONS

4.1 STUDY AIMS AND OBJECTIVES REVISITED

The aim of this thesis study was to spectrally decompose and describe the spatially regional patterns of savanna bush encroaching species across the Greater Kruger Park Region using the CAO integrated sensor data. This aim was achieved by making use of two main analytical approaches. Firstly, spectral similarity measures were utilised to understand the inter- and intra-species spectral separability between bush encroaching and savanna tree species. The significant bands and spectral regions in the hyperspectral data were also identified to assist the mapping procedure in the next approach. Secondly, a new novel decision tree multiple endmember SAM classification approach was introduced for the species level mapping in which 23 savanna tree species were classified and 4 bush encroaching species (the main focus) were mapped. For the objective, the species specific spectral and structural characteristics proved instrumental in the design of the decision tree and once the species were mapped, their distributions, across the various landscape features and land use practices, were clearly evident. The concluding remarks from the two analytical approaches will be elaborated in greater detail in the sub sections which will follow.

4.1.1 SAVANNA TREE AND BUSH ENCROACHING SPECIES DISCRIMINATION

Three spectral similarity measures; SAM, SID and SID-SAM (tan); were evaluated to determine which measure effectively discriminated the different savanna and bush encroaching species from one another. It was ascertained that the SAM measure yielded the highest probabilities of spectrally discriminating the majority of species under investigation in comparison to the other measures. SAM proved more effective especially for the spectrally similar species such as the bush encroaching species. Under the SAM results, it was evident that both the entire transformed spectrum and the primary and secondary plant chemical band scenarios yielded the bulk of the discriminating potential (highest SAM values) than the other spectral regions in the hyperspectral data. These results played a huge part in the mapping protocol in which a SAM hard pixel classifier was recommended, within a decision tree framework, with these significant bands being used to spectrally subset the hyperspectral data to improve mapping accuracies.

4.1.2 MAPPING OF SAVANNA BUSH ENCROACHING SPECIES

A decision tree multiple endmember SAM classification approach was considered as the preferred method for the mapping of savanna tree and bush encroaching species due to the shortcomings mentioned regarding alternative methods such as spectral unmixing. This new novel decision tree design, containing spectral and structural parameters, was tested against a traditional straightforward multiple endmember SAM classification approach to determine the highest overall classification accuracy obtained for classifying 23 tree species and for the mapping of 4 bush encroaching species. Various strategies were implemented to improve the classification accuracies yielded by the decision tree design. The finalised decision tree design which yielded the highest overall classification accuracy was obtained by a simple two end-node design which branched according to a 5 metre minimum tree height threshold. Subsetting the data according to the spectrally significant bands, from the previous chapter, was the most effective strategy which helped the final decision tree design reach an overall accuracy of 62.2% which surpassed the results of the traditional approach. With this final decision tree design, 4 bush encroaching species maps were made. From the bush encroaching species maps, 'encroached' areas, where these encroacher species were most concentrated, were identified as the entire Northern Sabi Sands region, along the riparian zones, on the mid-slopes and troughs landscape features to the east of Justicea and in the area to the north of Justicea. *Dichrostachys cinerea* was the most dominant encroaching species with the most prolific distribution while *Combretum apiculatum* was the least prevalent. The functionality of these map products depend on the management objectives of the different land use practices and the manner in which these bush encroaching species are perceived – a resource or a form of land degradation.

4.2 STUDY RECOMMENDATIONS

To promote opportunities for the continuation of future work and study regarding species level mapping in a complex and heterogeneous African savanna setting, various recommendations have been put forwards for the two main analytical chapters of this thesis

4.2.1 RECOMMENDATIONS FOR CHAPTER 2

It was recommended that the tree species be grouped together into family classes rather than multiple different species in an attempt to improve the discriminatory results. This was attempted but the results, however, was not favourable and did not lead to any improvements in the overall spectral separability between the species. The next step to improve results would be to streamline the spectral library of the 49 tree species by removing any species that do not have a good representative number of canopy level reference pixels (8-10 reference spectra should be the minimum target). Having insufficiently representative endmembers used in the classification process would lead to biased results in which there would be an over-representation of very limited species or under-representation of common species. As one of the pitfalls of the chapter study, intra-species spectral variability was not taken into account in the analyses because the mean reference spectrum, rather than the full range of training spectra, was used as reference endmember for each species (Cho et al, 2009). For this reason most of the *Combretum* and some *Acacia* species, in this study, yielded very low spectral separability values. The full range of training spectra for each species would encompass as much of the intra-species spectral variability due to differences in phenology, stress, climate and environmental conditions. Using this full range of training spectra for each species in these spectral similarity measure analyses would thus significantly improve the discriminatory ability between these species as it significantly improved the derived classification accuracies in the Cho et al., (2010) study. Majority of these recommendations were considered already in an attempt to bolster mapping efforts in chapter 3.

4.2.2 RECOMMENDATIONS FOR CHAPTER 3

It was recommended that alternative and more effective vegetation indices parameters be introduced in the decision tree design in order to assist the spectral component in playing a more active role in separating the species prior to classification. If a shortwave infrared (SWIR) spectral region was available for this study, indices which made use of this region would definitely produce more effective results in the overall decision tree architecture and possibly in the SAM classification as the spectral differences between the species would have been more pronounced. Scale is an important issue when viewing the species map products. If the user wanted to view the species maps at the canopy level, it was clear that a 'salt and pepper' mixed classification effect was evident in which multiple classes were designated to a single canopy which was unrealistic. Since this problem could not be resolved in this study,

more work needed to be done in order to develop an effective object orientated approach which recognised these tree canopies or crowns as individual entities for classification instead of a collection of pixels. If the user was more interested in viewing the map at a broader, regional scale of the entire study area, this mixed classification effect would not be a problem as the major species class would be clearly visible. Another recommendation to improve the species' accuracies, post classification, could be the implementation of topological rules for particular species depending on their likely locations in the landscape. For example, when classifying riparian zones, only the species that were typically found in these regions (based on field knowledge) should be subjected to classification while excluding the unlikely species. The same sort of procedure or rule set could be applied to landscape crests, mid-slopes and troughs etc. The automation of the decision tree technique in a particular software environment (e.g. Random Forest software), with an accompanying bootstrapping approach, would have provided a greater number of alternatives or possibilities for the optimum decision tree design and greater accuracies may have been reached. More quantitative statistics, regarding species density and richness, were recommended in order to make more solid conclusions regarding the influence of the different land use practices on the species dynamics across the study region as a whole. Other more complex, automated techniques such as the utilisation of Support Vector Machines could have also achieved higher classification accuracies but this type of techniques would need to be established further. Regardless of the limitations and proposed recommendations, this newly proposed method proved to be a pioneering attempt in the mapping of tree species in African savannas.

ACKNOWLEDGEMENTS

I would like to formally acknowledge and thank the following individuals who were instrumental in making this thesis study possible, enjoyable and worthwhile:

- My mentor Dr Moses Cho and my supervisor Dr Renaud Mathieu for their all contributions and wise guidance in helping me towards the completion of this MSc thesis study. Your knowledge of remote sensing and innovative methodologies were crucial in the development of this new novel approach introduced in this thesis.

- My supervisor Dr Frank Eckardt, in the Environmental and Geographical Science department of the University of Cape Town, for his advice on thesis structure, research inputs and for the associated proof reading.

- My work colleagues in the CSIR Ecosystems Earth Observation Unit for their warmth and willingness to assist me in the many problems, no matter how small, I've encountered along the way.

- Mr Kreelan Padayachee of the CSIR build environment who assisted in creating the Spectral Information Divergence script in Matlab and also assisted in fine tuning the Spectral Angle Mapper script needed for the analyses in the chapter 3.

- For the field data collected over the periods 2009 and 2010, I would like to acknowledge the following colleagues; Dr Moses Cho, Dr Renaud Mathieu, Mr Russell Main and Mr Abel Ramoelo; for their assistance in the study regions across Bushbuckridge, Sabi Sands Wildetuin and Kruger National Park.

- My friend Mr Michael Grover (assistant field ecologist) of Sabi Sands Wildetuin for hosting us and for the arrangement of game guards during the field visits.

- Dr Izak Smit of Kruger National Park Scientific Services for the arrangement of accommodation, access permits and game guards for the field visits in the Skukuza region.

- The numerous game guards, who accompanied us, guided us and protected us in the field visits in both Sabi Sands Wildetuin and Kruger National Park.

- Mrs Lesley Wright of CSIR who always assisted me in the arrangement of the bakkie hires and various fiscal and administrative tasks.

- The Andrew Mellon Foundation for the funding of the airborne remote sensing with the CAO. The CAO was made possible by the W.M Keck Foundation, the Gordon and Betty Moore Foundation and William Hearst III.

- Dr Greg Asner and his CAO unit in Stanford University, USA, for arranging the flight campaign of the CAO sensor in 2008.

- T. Kennedy-Bowdoin, D. Knapp, J. Jacobson, and R. Emerson at the Carnegie Institution for Science for the hyperspectral and LiDAR data pre-processing support from which the data products used in this thesis were derived.

REFERENCES

- Affendi. S, N.A Ainuddin & H.Z.M Shafri (2009). A rule based approach for the mapping of tropical forest canopy from airborne hyperspectral data. *GIS development.net Online journal*; pp 1-13

- Asner. G.P (2008 –c). Hyperspectral remote sensing of canopy chemistry, physiology and diversity in tropical rainforests. Chapter 12 in Hyperspectral remote sensing of tropical and subtropical forests. M. Kalacska and G.A. Sanchez-Azofeifa (eds.) Taylor and Francis

- Asner. G.P & K.B Heidebrecht (2002). Spectrally unmixing of vegetation, soil & dry carbon cover in arid regions: comparing multispectral & hyperspectral observations. *International Journal of Remote Sensing*. Vol 23; Issue 19; pp 3939-3958

- Asner, G.P., M.O. Jones, R.E. Martin, D.E. Knapp, and R.F. Hughes (2008 – a). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment*. Vol 112; pp 1912-1926

- Asner. G.P, D.E Knapp, T. Kennedy-Bowdoin, M.O Jones, R.E Martin, J. Boardman & C.B Field (2007). Carnegie Airborne Observatory: in-flight fusion of hyperspectral imaging and waveform LiDAR for 3D studies of ecosystems. *Journal of Applied Remote Sensing*. Vol 1; pp 1-27

- Asner. G.P, D.E Knapp, T. Kennedy-Bowdoin, M.O Jones, R.E Martin, J. Boardman & R.F Hughes (2008 - b). Invasive species detection in Hawaiian Rainforests using In-flight fusion of Airborne Imaging Spectroscopy & LiDAR. Carnegie Institution: Stanford University; pp 1-33

- Asner. G.P & T.A Varga (2008). Hyperspectral & LiDAR Remote Sensing of Fire Fuels in Hawaiian Volcanoes National Park. Carnegie Institute: Stanford University; pp 1-34

- Atkinson. P.M & N.J Tate (editors) (1999). Advances in remote sensing and GIS analysis. *John Wiley & Sons; Chichester, West Sussex, England*. Chapter 2 – Land cover classification revisited by P.M Mather; pp 7-16

- Bai. Z.A, D.L Dent, L. Olsson & M.E Schaepman (2008). Global assessment of land degradation and improvement – 1. Identification by remote sensing. *GLADA Report*. Issue 5; pp 1-78

- Baldi. G & J.M Paruelo (2008). Land use and land cover dynamics in South American, temperate grasslands. *Ecology and Society*. Vol 13; Issue 2; Article 6; pp 1-20

- Ben-Shahar. R (1991). Abundance of trees and grasses in a woodland savanna in relation to environmental factors. *Journal of Vegetation Science*; Vol 2; Issue 3; pp 345-350.

- Ben-Shahar. R (1992). The effects of bush clearance on African ungulates in a semi-arid nature reserve. *Ecological Applications*; Vol 2; Issue 1; pp 95-101

- Bork. E.W & J.G Su (2007). Integrating LiDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment*. Vol 111; pp 11-24

- Boschetti. M, L. Boschetti, S. Oliveri, L. Casati & I. Canova (2007). Tree species mapping with airborne hyperspectral MIVIS data: the Ticino Park case study. *International Journal of Remote Sensing*. Vol 28; Issue 6; pp 1251-1261

- Chambers. J.Q, G.P Asner, D.C Morton, L.O Anderson, S.S Saatchi, F.D.B Espirito-Santo, M. Palace & C. Souza Jr (2007). Regional ecosystem structure & function: ecological insights from remote sensing of tropical rainforests. *Trends in Ecology & Evolution*. Vol 22; Issue 8; pp 414-424

- Cho. M.A, P. Debba, R. Mathieu, J. van Aardt, G.P Asner, L. Naidoo, R. Main, A. Ramoelo & B. Majeke (2009). Spectral variability within species and its effects on savanna tree species discrimination. *IEEE International Geoscience & Remote Sensing Symposium (IGARSS) 2009 Conference Paper*, University of Cape Town, July 2009

- Cho. M.A, P. Debba, R. Mathieu, L. Naidoo, J. van Aardt & G.P Asner (2010). Improving discrimination of savanna tree species through a multiple endmember spectral angle mapper (SAM) approach: canopy level analysis. *IEEE International Journal of Geoscience & Remote Sensing* – Status: submitted

- Clark. M.L, D.A Roberts, D.B Clark (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf and crown scales. *Remote Sensing of Environment*. Vol 96; pp 375-398

- Cortijo. F.J & N. Perez de la Blanca (1996). Image classification using non-parametric classifiers and contextual information. *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI, Part B3. Vienna 1996; pp 120-124

- Dalle. G, B.L. Maass & J. Isselstein (2006). Encroachment of woody plants and its impact on pastoral livestock production in the Borana lowlands, southern Oromia, Ethiopia. *African Journal of Ecology – East African Wildlife Society*. 44, pp 237-246

- Du. Y, C-I Chang, H. Ren, C-Chi Chang, J.O Jensen & F.M D'Amico (2004). New hyperspectral discrimination measure for spectral characterization. *Society of Photo-Optical Instrumentation Engineers*. Issue 43; Vol 8; pp 1777-1786

- Eastman. J.R (2006). IDRISI Andes: Guide to GIS and Image Processing (Manual Version 15). *Clark Labs, Clark University*; Worcester, MA, USA

- Eckhardt. H.C & H.C. Biggs (In. press). Medium and long term changes in woody vegetation cover in Kruger National Park, South Africa. *South African National Parks*, pp 145-148

- ENVI version 3.2 User Guide (July 1999 Edition). *Online Manual*; ITT Visual Information Systems

- Eswaran, H., R. Lal and P.F. Reich (2001). Land degradation: an overview. Responses to Land Degradation. Proc. 2nd. International Conference on Land Degradation and Desertification, Khon Kaen, Thailand. Oxford Press, New Delhi, India.

- Fisher. J, B.F.N. Erasmus, E. Witkowski, J. van Aardt, G. Asner, T. Kennedy-Bowdoin, D. Knapp, R. Mathieu, and K. Wessels (2009). Three-dimensional woody vegetation structure across different land-use types and land-use intensities in a semi-arid savanna. Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS-2009), Cape Town, South Africa.

- Gamon. J, J. Penuelas & C.B Field (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of environment*. Vol 41; pp 35-44.

- Geerling. G.W, M. Labrador-Garcia, J.G.P.W Clevers, A.M.J Ragas & A.J.M Smits (In. press). Classification of floodplain vegetation by data-fusion of Spectral (CASI) and LiDAR data. *International Journal of Remote Sensing*. Research Article; pp 1-24

- Giannecchini. M, W. Twine & C. Vogel (2007). Land-cover change and human-environment interactions in a rural cultural landscape in South Africa. *The Geographical Journal*. Vol 173; Issue 1; pp 26-42

- Gitelson, A.A & M.N. Merzlyak (1994). Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology*. Vol 143; pp 286-292.

- Gitelson, A.A, Y. Zur, O.B. Chivkunova & M.N. Merzlyak (2002). Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochemistry and Photobiology*. Vol 75; pp 272-281.

- Hestir. E.L, S. Khanna, M.E Andrew, M.J Santos, J.H Viers, J.A Greenberg, S.S Rajapakse & S.L Ustin (2008). Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Remote Sensing of Environment*. Vol 112; pp 4034-4047

- Hudak. A.T & C.A. Wessman (2001). Textural Analysis of high resolution imagery to quantify bush encroachment in Madikwe Game Reserve, South Africa, 1955-1996. *International Journal of Remote Sensing*. Vol 22; Issue 14; pp 2731-2740

- Hudak. A. T & C.A Wessman (1998). Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna. *Remote Sensing of Environment.* Vol 66; pp 317-330.

- Jafari. R, M.M Lewis & B. Ostendorf (2006). The use of EO-1 hyperspectral imagery for the discrimination of arid vegetation. Proceeding of the 13th Australian Remote

Sensing & Photogrammetry Conference. The Photogrammetry Association of Australia.

- Joseph. K.A (2005). Comparison of segment and pixel based non-parametric classification of land cover in the Amazon region of Brazil using multi-temporal Landsat TM/ETM+ Imagery. *MSc thesis – Virginia Polytechnic Institute & State University, USA*; pp 1-153

- Joubert, S (2007). The Kruger National Park – A history (Volumes I, II & III), 1st edition; *High Branching (Pty) Ltd*, Johannesburg: South Africa.

- Kim. C (1996). A winter cover classification of lower Nakdong River region using JERS-1 OPS data. *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI; Part B7. Vienna; pp 341-346

- Kim. S (2007). Individual tree species identification using LiDAR-derived crown structures and intensity data. *Doctoral thesis*. University of Washington, College of Forest Resources; pp 1-137

- Kooistra. L, L. Sanchez-Prieto, H.M Bartholomeus, M.E Schaepman (In. Press). Regional mapping of plant functional types in river floodplain ecosystems using airborne imaging spectroscopy data. *Commission VII*; pp1-6

- Landmann. T (In. press). A case study for Skukuza: Estimating biophysical properties of fires using EOS-MODIS satellite data. *Unpublished thesis*; pp 1-179.

- Lawrence. R.L & W.J Ripple (1998). Comparisons among vegetation indices and bandwise regression in a highly disturbed heterogeneous landscape: Mount St. Helens, Washington. *Remote Sensing of Environment*. Vol 64; pp 91-102

- Lees. B.G & K. Ritman (1991). Decision tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management*. Vol 15; Issue 6; pp 823-831

- Lefsky. M.A, W.B Cohen, G.G Parker & D.J Harding (2002). LiDAR remote sensing for ecosystem studies. *Bioscience Articles*. Vol. 52; Issue 1; pp 19-30

- Lieng. E, D. Vikhamar Schuler, L. Kastdalen, G. Fjone, M. Hansen & J.P Bolstad (In. press). Classification of land cover using decision trees and multiple reference data sources. *Online journal*; pp 1-4

- Lilliesand. T.M & R.W. Kiefer (1994). Remote Sensing and Image Intepretation (3rd edition), *John Wiley & Sons, Inc*, New York: USA.

- Lyon. J.G, D. Yuan, R.S Lunetta & C.D Elvidge (1998). A change detection experiment using vegetation indices. *Photogrammetric Engineering & Remote Sensing*. Vol 64; Issue 2; pp 143-150

- Martin, R.E & G.P Asner (2005). Regional estimation of Nitric Oxide emissions following woody encroachment: linking image spectroscopy and field studies. *Ecosystems*. Issue 8; pp 33-47.

- McGlynn. I.O & G.S Okin (2006). Characterisation of shrub distribution using high spatial resolution remote sensing: Ecosystem implications for a former Chihuahuan Desert grassland. *Remote Sensing of the Environment*. Issue 101; pp 554-566

- Naidoo. L (2007). Quantification of bush encroachment using aerial photography of the Pretoriuskop region, in the Kruger National Park, from the period 1965 to 2003. BSc Honours thesis – unpublished; Department of Environmental and Geographical Science, University of Cape Town.

- Odagawa. S & K. Okada (In. press). Tree species discrimination using continuum removed airborne hyperspectral data. *Earth Remote Sensing Data Analysis Center*. Preliminary draft; pp 1-4

- Oldeland. J, W. Dorigo, D. Wesuls & N. Jürgens (2010). Mapping bush encroaching species by seasonal differences in hyperspectral imagery. *Open Access: Remote Sensing*. Vol 2; pp 1416-1438

- Pouncey. R, K. Swanson & K. Hart (1999). ERDAS Field Guide (5th Edition, Revised and Expanded). *ERDAS, Inc. Atlanta, Georgia, USA*; pp 1-698

- Pu. R, P. Gong, Y. Tian, X. Miao, R.I Carruthers & G.L Anderson (2008). Using classification and NDVI differencing methods for monitoring sparse vegetation

coverage: a case study of saltcedar in Nevada, USA. *International Journal of Remote Sensing*. Vol 29; Issue 14; pp 3987-4011

- Richards. J.A & X. Jia (2005). Remote sensing digital image analysis: an introduction (4th Edition). *Springer Berlin Heidelberg New York*. Chapter 3 – The interpretation of Digital Image Data; pp 67-81

- Roberts. D.A, P. Dennison, S.L Ustin, E. Reith & M. Morais (In press). Development of a regionally specific library for the Santa Monica Mountains using high resolution AVIRIS data. In: Proceedings 8th AVIRIS Earth Science Workshop (p.6), Pasadena, CA: JPL; pp 1-7

- Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering (1973). Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351 I; pp 309-317.

- Schmidt. E, M. Lotter & W. McCleland (2007). Trees and shrubs of Mpumalanga and Kruger National Park (2nd Edition). *Jacana Media*, Johannesburg, South Africa

- Schowengerdt. R.A (1983). Techniques for image processing and classifications in remote sensing (1st Edition). *Academic Press*. Chapter 3 – Digital image classification.

- Shackleton. C.M & S. Shackleton (2003). Value of non-timber forest products and rural safety nets in South Africa. *Paper presented at the International Conference on Rural Livelihoods, Forests and Biodiversity*; 19-23 May, Bonn, Germany; pp 1-18

- Shackleton. C.M (2000). Comparison of plant diversity in protected and communal lands in Bushbuckridge lowveld savanna, South Africa. *Biological Conservation.* Issue 94; pp 273-285.

- Shackleton. C.M, G. Guthrie & R. Main (2005). Estimating the potential role of commercial over-harvesting in resource viability: A case study of five useful tree species in South Africa. *Land degradation & Development.* Vol 16; pp 273-286

- Shafri. H.Z.M, S. Affendi & S. Mansor (2007). The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in

hyperspectral image analysis. *Journal of Computer Science*. Vol 3. Issue 6; pp 419-423

- Soban. Md.I (2007 - a). Species discrimination from a hyperspectral perspective. *Doctoral Thesis* – International Institute for Geo-information Science & Earth Observation, Enschede, the Netherlands. Chapter 1 – General introduction & Chapter 4 – Spectral similarity measures for plant species discrimination; pp 61-80

- Soban. Md.I (2007 – b). Species discrimination from a hyperspectral perspective. *Doctoral Thesis* – International Institute for Geo-information Science & Earth Observation, Enschede, the Netherlands. Chapter 6 – Mapping shrub and tree species richness from hyperspectral imagery using a matched filtering unmixing technique; pp 103-124

- Sun. F, S. Omachi, N. Kato, H. Aso, S. Kono & T. Takagi (2000). Two-stage computational cost reduction algorithm based on Mahalanobis distance approximations. *International Journal of Pattern Recognition*. Vol 2; pp 2969-2973

- Thenkabail. P.S, R.B Smith & E. De Pauw (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*. Vol 71; pp 158-182

- Tong. Q, B. Zhang & L. Zheng (2004). Hyperspectral remote sensing technology and applications in China. *Proceedings of the 2nd CHRIS/Proba Workshop, ESA/ESRIN*, Frascati, Italy, 28-30 April; pp 1-10

- Ustin. S.L, D.A Roberts, J.A Gamon, G.P Asner & R.O Green (2004). Using imaging spectroscopy to study ecosystem processes and properties. *Bioscience – Articles*. Vol 54, Issue 6; pp 523-534

- Vaiphasa. C, S. Ongsomwang, T. Vaiphasa & A.D Skidmore (2005). Tropical mangrove species discrimination using hyperspectral data: A laboratory study. *Estuarine, Coastal and Shelf Science*. Vol 65; pp 371-379

- Van der Meer. F (2006). The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*; pp 1-15

- Ward. D (2005). Do we understand the causes of bush encroachment in African savannas? *African Journal of Range and Forage Science*, 22(2), pp 101-105

- Wessels. K.J, S.D. Prince, P.E. Frost & D. Van Zyl (2004). Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1km AVHRR NDVI time series. *Remote Sensing of Environment*. 91, pp 47-67

- Wigley. B.J, W.J Bond & M.T Hoffman (2009). Bush encroachment under three contrasting land-use practices in a mesic South African savanna. *African Journal of Ecology*. Vol 47; tissue 1; pp 62-70

- Yingchun. S, Z. Xianfeng, C. Xiuwan, H. Zhaoqiang & W. Caicong (2006). Mangrove type classification using airborne hyperspectral images at Futian reservation, Shenzhen, China. *Geoscience and Remote Sensing Symposium 2006, IGARSS 2006, IEEE International Conference*; pp 3451-3454

APPENDICES

1A. CAO SYSTEM CONFIGURATION

The CAO sensor was flown as a collaborative research effort between the Council for Scientific and Industrial Research (CSIR), South African National Parks and the CAO unit of Stanford University in the US. A mid-altitude aircraft, which was equipped with this integrated hyperspectral and LiDAR system, was flown in conjunction with the May 2008 field campaign in which the study regions were sampled on the ground to minimise temporal differences and associated errors between the ground-truthing and image datasets.

The hyperspectral scanner of the airborne integrated sensor system, which was used during the field campaign, employed a pushbroom imaging array with 1500 cross-track pixels. It possessed a spectral range of 0.4-1.2μm (covering the visible and Near-Infrared portions of the spectrum) with a spectral resolution or wavelength interval of approximately 9.23nm. The Shortwave Infrared (SWIR), which was considered vital by the global remote sensing community for most in-depth vegetation related studies, was not utilized at the time of the campaign as the signal-to-noise ratio was too low for effective operability (Asner, personal communication – April 2008). The hyperspectral scanner also possessed two spatial resolution settings of 1m (low resolution) and 0.5m (high resolution) which allows highly detailed imagery to be created (each pixel on the image represents a 0.5m by 0.5m or 1m by 1m area on the ground). A waveform footprint type LiDAR instrument was used in this integrated system. As with the hyperspectral sensor, high resolution & low resolution LiDAR datasets were also created. The 1m spatial resolution datasets were utilized for this thesis study as it covered a greater spatial extent to observe the ecosystem under question. The LiDAR differs from passive sensors in that it makes use of a laser altimeter which actively emits highly repeating laser pulses of a given size or footprint. These highly repeating laser pulses have multiple returns that accurately measure the distance from the sensor to the canopy top and bottom topography (Chambers et al, 2007). The two instruments were coupled with an onboard Global Positioning System-Inertial Measurement Unit (GPS-IMU) to allow for the seamless co-location, geo-referencing and orthorectification of the two data types in three-dimensional space during flight. In essence, the GPS-IMU data form the common link for the detailed ray tracing of the photons between aircraft sensors and the ground. Ultimately, the CAO system collects at least one LiDAR pulse for every

hyperspectral pixel. The technical specifications of this CAO integrated sensor is summarized in the table below followed by an image describing the sensor's configuration:

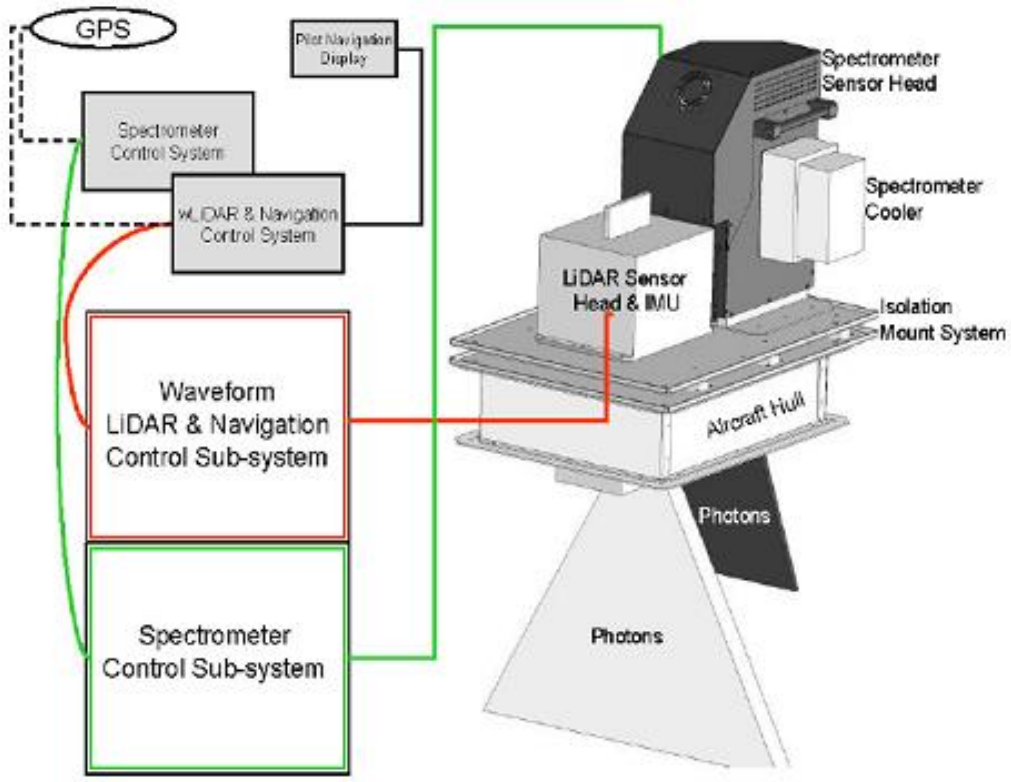


Figure 1A: The CAO integrated sensor platform (Asner et al, 2007)

1B. CAO SENSOR SYSTEM SPECIFICATIONS (SUMMARIZED)

CAO Sensor System Specifications					
Platform	Origin	Number of Bands	Spectral Range	Spectral interval/resolution	Spatial Resolution
High-Fidelity Image Spectrometer (HiFIS)	Stanford University, California, United States	72	384.8-1054.3 nm	9.23nm	1.12m
Waveform Light Detection & Ranging (LiDAR)		N/A	Operates at 1064nm	Laser pulse repetition rate of 50KHz	0.56m to 1.12m (depending on flight altitude)

2A. SPECTRAL SIMILARITY MEASURE FORMULAE

$$SAM(sp1;sp2) = cos^{-1}\left(\frac{\sum_{l=1}^{L} sp1l \times sp2l}{[\sum_{l=1}^{L} sp1^2]^{\frac{1}{2}}[\sum_{l=1}^{L} sp2^2]^{\frac{1}{2}}}\right) \quad (1)$$

$$SID\ (sp1;sp2) = D(sp1 \parallel sp2) + D(sp2 \parallel sp1) \quad (2)$$

$$where\ D(sp2 \parallel sp1) = \sum_{l=1}^{L} ql[Il(sp1) - Il(sp2)]$$

$$and\ D(sp1 \parallel sp2) = \sum_{l=1}^{L} pl[Il(sp2) - Il(sp1)]$$

$$\text{SID-SAM(tan)} = \text{SID(si, sj)} \times \tan[\text{SAM(si, sj)}] \quad (3)$$

$$PSD\ \ Pt, \Delta(k) = \frac{m(t, sp1)}{\sum_{j=1}^{L} m(t, sp2)}, \quad for\ k = 1, \ldots . K \quad (4)$$

2B. PLANT CHEMICAL COMPONENTS AND THEIR ASSOCIATED WAVELENGTHS

Plant Chemical Component	Wavelengths (nm)
Chlorophyll a	420; 490; 660
Chlorophyll b	435; 643
β Carotene	425; 450; 480
α Carotene	420; 440; 470
Xanthophylls	425; 450; 475
Starch	990;
Oil	930; 1040
Proteins	910; 1020

2C. 49 SAVANNA TREE SPECIES LIST AND ACCOMPANYING ABBREVIATIONS

Abbreviations	Tree Species Names	# of Spectra
A.squasensis	A. squasensis	6
Acacia_exuv	Acacia exuviales	7
Acacia_gerr	Acacia gerrardii	50
Acacia_grandi	Acacia grandicomuta	7
Acacia_nigres	Acacia nigrescens	64
Acacia_sp1	Unidentified Acacia species 1	5
Acacia_sp2	Unidentified Acacia species 2	5
Albiz_versi	Albizia versicolor	19
Combr_apic	Combretum apiculatum	27
Combr_coll	Combretum collinum	71

Combr_here	Combretum hereroense	33
Combr_imber	Combretum imberbe	91
Combr_molle	Combretum molle	8
Combr_zey	Combretum zeyheri	53
Combr_sp1	Unidentified Combretum species 1	6
Cross_pterex	Crossopterex	30
Dichro_ciner	Dichrostachys cinerea	34
Domb_rotun	Dombeya rotundifolius	6
Euclea_divi	Euclea divinorum	20
Euclea_nat	Euclea natalensis	38
Fig	Fig	13
Flueg_viro	Flueggea verosa	14
Gimno_bux+sene	Mixed G.buxifolia & G.senegalensis	51
Gimno_bux	Gimnosporia buxifolia	10
Gimno_sene	Gimnosporia senegalensis	41
Greenthorn	Greethorn Tree	18
Grewia_flav	Grewia flavescens	9
Grewia_sp1	Unidentified Grewia species 1	6
Lannea_disco	Lannea discolor	23
Loncho_cap	Lonchocarpus capassa	36
L9T7	Unknown tree species (tree #7 in transect 9)	4
Ozor_pani	Ozoroa paniculosa	15
P.gardeniifolia	P. gardeniifolia	7
Pelto_afri	Peltophorum africanum	16
Pilio_thonn	Piliostigma thonningii	25
Ptero_rotun	Pterocarpus rotundifolius	48
Ptero_angol	Pterocarpus angolensis	9
Rhus_sp1	Unidentified Rhus species 1	9
Rhus_sp2	Unidentified Rhus species 2	5
Sclero_birr	Sclerocarya birrea	192
Strych_madag	Strychnos madagascariensis	8
Strych_mozam	Strychnos mozambiquensis	8
Strych_sp1	Unidentified Strychnos species 1	4
T22	Unknown tree species (tree #22, unknown transect)	3
Term_seri	Terminalia sericea	71
Unknown401	Unknown tree species 1	4
Unknown601	Unknown tree species 2	3
UnknownL9920	Unknown tree species 3	5
Zizi_macro	Ziziphus mucronata	21

2D. SAM MATLAB SCRIPT

```
function SAM=sam(ref)

x=length(ref(1,:));
A=ref.^2;
B = zeros(size(ref,1),x,x);
for i=1:x
    for j=1:x
        B(:,i,j)=ref(:,i).*ref(:,j);
    end
end
C=sqrt(sum(A));
D=reshape(sum(B),[x,x]);
for m=1:x;
  for n=1:x
E(m,n)=C(:,m).*C(:,n);
  end
end
SAM=acos(D./E);
```

2E. SID MATLAB SCRIPT

```
clc;
clear all;

format long
[raw,names] = xlsread('SID_EntireSpectrum2.xlsx','Mean' );

sizem = size(raw);
row = sizem(1);
col = sizem(2);
np = col-2;

for i=1:1:row
    for j=1:1:np
    param(i,j)=raw(i,j+2);
    end
end

for i=1:1:np
    p(i).sum = sum(param(:,i));
end

for i=1:1:np
  for j=1:1:row
    p(i).pk(j) = param(j,i)/p(i).sum;
    p(i).ln(j) = -log(p(i).pk(j));
  end
end

for i=1:1:np
    for j=i+1:1:np
        for k=1:1:row
            p(i).comp1(j,k) = p(i).pk(k)*(p(j).ln(k)-p(i).ln(k));
            p(i).comp2(j,k) = p(j).pk(k)*(p(i).ln(k)-p(j).ln(k));
        end
```

```
    p(i).SID(j) = sum(p(i).comp1(j,:))+sum(p(i).comp2(j,:));
    end
end

for i=1:1:np-1
    for j=1:1:np
        outputm(i,j) = p(i).SID(j);
    end
end

xlswrite('SID_EntireSpectrum2.xlsx', outputm, 'Output', 'A1');
```

3A. 23 SAVANNA TREE SPECIES INFORMATION: CODES UTILIZED IN THE RESULTS SECTION AND ATTRIBUTE INFORMATION (SCHMIDT ET AL, 2007 & SHACKELTON ET AL, 2005)

Scientific Name	Common Name	Code	Attributes
Acacia gerrardii	Red Thorn	AG	Shrub to medium sized tree. Erect branches and a flattened crown Bark is grey to blackish and rough. Younger branches are reddish and hairy Thorns are in short pairs. Leaves are tiny and clustered on prominent woody cushions. Fruit are sickle-shaped, hairy pods. Thorny bush encroaching species. Bark contains constricting tannin chemicals used for medicinal purposes and the inner bark is used to create twine.
Acacia nigrescens	Knob Thorn	AN	Medium to large tree up to 30m. Common in arid bushveld across soil types Bark is brown to black and covered with persistent thorn-tipped knobs Thorns are in hooked pairs and almost black. Leaves are twice-compound, leathery and hairless. Fruit are straight, olive to black pods.
Berchemia discolor	Brown Ivory	BD	Generally large tree up to 20m. Usually on river banks and on termitaria Pale green covered in brown lenticels when young. Bark is dark grey and roughly fissured. Leaves are simple and slightly ovate or elliptic. Side veins form a distinctive herringbone pattern.
Combretum apiculatum	Red Bushwillow	CA	Small to medium sized tree. Widespread across savanna. Mostly single stemmed Bark is dark blackish or brownish grey. Leaves are oblong or broadly ovate. Leaf apex has a characteristic twisted tip. Dull to glossy green and paler below. Very common savanna tree species which is highly abundant in the study region. Good for charcoal production and possesses numerous medicinal properties (treats certain snakebite and dysentery).
Combretum collinum	Weeping Bushwillow	CC	Medium-sized, multi-stemmed tree. Bark is pale greyish brown and rough Drooping pale green leaves with dense woolly hairs when young Flowers are small and produced in long spikes
Combretum hereroense	Russet Bushwillow	CH	Small sized tree. Single stemmed with pale to blackish grey bark. Leaves are oval with a rounded apex with a short, twisted point. Fruits are 4-winged have a

			distinct colouring pattern (reddish brown body with golden brown wings)
Combretum imberbe	Leadwood	CI	Medium to large tree with a spreading canopy. Found on alluvial or brackish/ basaltic soils. Large upright stem with pale grey bark. Bark is rough with distinct cracking into deep vertical and horizontal fissures. Leaves are distinctly pale beneath.
Combretum zeyheri	Large-fruit Bushwillow	CZ	Small to medium sized tree with sparse rounded crown. Single or multi-stemmed with low down branching. Older stems are pale to very pale grey with darker grey patches. Leaves are elliptic to oblong. Fruit are 4-winged and distinctly large (60 X 60mm)
Dalbergia melanoxylon	Zebrawood	DalMel	Large woody shrub or small multi-stemmed tree. Forming large stands on clay soils. Bark is smooth and pale grey. Branches are modified to form spines. Leaves are compound and dark green above.
Dichrostachys cinerea	Sickle-bush	DC	Shrub or small rounded tree, often encroaching if veld is mismanaged. Branching low down and bark is rough with fissures. Side twigs are modified to form spines. Small leaves clustered on spines/side shoots. Fruit occurs distinctively as a curled and twisted mass of brown pods. Hardy and pervasive bush encroaching species which impedes cattle and local movements.
Diospyros mespiliformis	Jackal-berry	DM	Large tree (15-20m) often associated with termite mounds and riverine fringes Leaves are elongated and dark glossy green above. Trees rarely appear bare Flowers are small and fruits are edible, round and yellowish when ripe
Euclea natalensis	Hairy Guarri	EN	Shrub to medium sized tree. Usually associated with dense thickets or riverine areas. Single or multi-stemmed. Branches are slender, crooked and greyish brown. Bark is grey to blackish. Leaf margin often wavy, glossy dark green to matt

			olive green. Fruit are small, round and thinly fleshed (red ripening to black)
Gimnosporia senegalensis	Red Spikethorn	GS	Shrub or small often multi-stemmed tree. Usually hanging to one side and common along rivers and dry riverbeds. The main stem is reddish when young becoming greyish-green. Spines to 70mm, sometimes bearing leaves to form small spine-tipped branches. Leaves are clustered, oblong. Midrib is much paler than the rest of the leaf.
Lannea discolor	Live-long	LD	Smallish tree with a neat, round crown. Often found on steep slopes and rocky outcrops. Bark is dark grey, roughish and cracking into blocks. Branchlet tips thickened with prominent leaf scars.
Lonchocarpus capassa	Apple Leaf	LC	Currently known as Philenoptera violacea. Rounded tree up to 18m. Usually found on alluvial soils, close to rivers. Compound leaves with three large leaflets. Leaves are greyish and slightly hairy. Bark is light grey. Flowers are bright blue. Fruit is a narrow flat pod (greyish brown).
Peltophorum africanum	African-wattle	PA	Small to medium sized rounded tree. Frequently branched near the base. Brown bark which is rough and fissured. Young branches are densely covered in rusty-brown hairs. Leaves are dull olive-green above and paler below. Flowers are large yellow sprays
Pterocarpus rotundifolius	Round-leaved Bloodwood	PR	Large, rounded, woody shrub or tree. Often forming dense colonies. Usually multi- stemmed with grey young bark. Leaflets are large and rounded with distinguishing parallel side veins. Active bush encroaching species. Good for apiculture due to the rich pollen and nectar sources and plays a role in soil erosion control.
Schotia brachypelata	Weeping Boer-bean	SchotB	Medium to large sized tree with rounded canopy. Bark is grey becoming brown. Leaves are compound with 4-7 pairs of opposite leaflets. Flowers are clustered in dense red sprays. Petals reduced to linear filaments.

Sclerocarya birrea	Marula	SB	Common in SA savannas especially on sandy frost free soils Large and dominant tree (up to 20m). Protected tree species in SA Leaves are compound, dark green above and paler below Separate male and female trees. Has a large ovoid tasty fruit
Spirostachys africana	Tamboti	SA	Large erect tree with round canopy and common on brackish flats and along seasonal streams and rivers. Occur in dense stands. Bark is very dark with cracks in rectangular blocks. White latex is present. Leaves are simple and ovate. Have small glands present on top of the petiole at the base. Fruit is a 3-lobed capsule with brown seeds
Strychnos sp.	Monkey Orange	Stry	Small shrubby to medium sized tree. Single or multi-stemmed. Densely branched with pale grey bark (smooth). Short rigid lateral shoots emerge from the branches. Fruit are very large and round with a thick woody shell (bluish-green to yellowish- orange when ripe)
Terminalia sericea	Silver Cluster-leaf	TS	Small to medium sized tree with rounded crown to characteristically flat-topped Upright stem with reddish-brown to purplish-brown branches. Often bearing small rounded woody galls. Leaves are crowded at the branch ends. Foliage have a distinct blue-grey colour at a distance. Although being a known bush encroaching species, it is primarily utilised as fuel wood to satisfy the energy requirements of local communities
Ziziphus mucronata	Buffalo-thorn	ZM	Medium-sized with most species between 4-6m. Dense in areas where ground water is readily available. Dark, shiny leaves with three veins clearly visible radiating from the asymmetrical base. Leaf margin is finely serrated. Armed with paired spines. Round, small, stony fruit (reddish-brown when ripe).

3B. TOTAL NUMBER OF SAMPLED PIXELS, PER SPECIES, FOR THE TRAINING AND VALIDATION DATASETS

L456	**Training**	**Validation**
Species	Total # of pixels	Total # of pixels
AG_DC	76	77
AN	251	250
BD	28	28
CA	100	100
CC	41	41
CH	32	31
CI	70	69
CZ	18	17
DalMel	14	14
DC	145	144
DM	31	31
EN	25	25
GS	19	19
LD	13	12
LC	32	31
PA	27	26
PR	50	49
SchotB	16	15
SB	275	274
SA	82	81
Stry	9	9
TS	55	54
ZM	47	47

3C. CONFUSION MATRIX – MINOR PRUNING STRATEGY

Minor Pruning (4 End Nodes)	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	46	2		1		1	4		2	4		2				2	1	1				1	2	69	66.7
AN		180		1	2	4	10		6	6				4		2	1		25			3	3	247	72.9
BD		1	14								1			2				7	3					28	50.0
CA	12			51	3	4	9	3		9		1	1			2	1					2	2	100	51.0
CC	3	3		1	6	3	2		2	2	1	3		1		2	3	4	4				1	41	14.6
CH	2	2				18	4	1		3													1	31	58.1
CI	2	5					35	2					2			1		17	5					69	50.7
CZ	1						2	11		1								1				1		17	64.7
DalMel				1					7	2					1		1						2	14	50.0
DC	17	2		13	6	13	9	2		38		3	2			8	8		2			2	9	134	28.4
DM			1								14							12	4					31	45.2
EN	1						2			2		17				2						1		25	68.0
GS	1				1	2		1		3		2	6			2								18	33.3
LC	4	2				1	3			1	2	1		8				5	3			1		31	25.8
LD						1		1				1			7			2						12	58.3
PA	5	2		1		1	2			2			1			9	1	1					1	26	34.6
PR	3	2		2	1	6	3	2	1	4		1				4	12		1			1		43	27.9
SA		1	1								1			2				70	4	2				81	86.4
SB		33	1	1			15							7		1		14	199	2		1		274	72.6
SchotB																		4		11				15	73.3
Stry																		1		3	5			9	55.6
TS	3			3	1	3	3			8		1				1	2		1			26	1	53	49.1
ZM	4	2		2		4	6		2	3		1	1	1		2		2	2			3	7	42	16.7
Grand Total	**104**	**237**	**17**	**77**	**20**	**61**	**109**	**23**	**20**	**88**	**19**	**33**	**13**	**25**	**8**	**38**	**30**	**141**	**253**	**18**	**5**	**42**	**29**	**1410**	
Producer's Accuracy	44.2	75.9	82.4	66.2	30.0	29.5	32.1	47.8	35.0	43.2	73.7	51.5	46.2	32.0	87.5	23.7	40.0	49.6	78.7	61.1	100.0	61.9	24.1		
Overall Accuracy	56.53																								

3D. CONFUSION MATRIX – MINOR PRUNING STRATEGY + SELECTED TRAINING ENDMEMBERS

Minor Pruning + Selected Endmembers	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	42	1		1	1	2	5		2	3			2			1	3		2				2	67	62.7
AN		173		1	2	1	16		6	5				7		1	1	1	20			4	5	243	71.2
BD			11								2			3				9	2		1			28	39.3
CA	6			42	6	7	9	2		10		1	1			5	2					2	2	95	44.2
CC	1	4		1	5	4	5		2	1	1	4		2		2	3	5					1	41	12.2
CH	2	1		2		9	7	1		3			1				2						3	31	29.0
CI	3	5	1				30	1		2			3	2		1		14	7					69	43.5
CZ	2			1			2	8		1						2				1				17	47.1
DalMel				1			1		7	2						1	1						1	14	50.0
DC	10	2		10	4	16	12	2		30		6	2			11	13	2				2	7	129	23.3
DM			3				1				13			1				12	1					31	41.9
EN						1	3			2		15	1			3								25	60.0
GS	1				1	2	2			3		1	6			2								18	33.3
LC	2	3				1	3			2	2	1		8		1		5	2					30	26.7
LD					1							1	1		6	1		2						12	50.0
PA	4	2		1			4		1	2				1		9	2							26	34.6
PR	2	3		3		2	3			3		1	2			3	15		1			2	1	41	36.6
SA		2	1				2				2			2				66	4	2				81	81.5
SB		57		1		1	23				6			14		1		18	150	2		1		274	54.7
SchotB			1															4		10				15	66.7
Stry																		1		3	5			9	55.6
TS	2			1	1	4	3			5		1				1	3		1			25	7	54	46.3
ZM	3	1		1		4	5		3	3		1		1		2		2	3			4	6	39	15.4
Grand Total	**80**	**254**	**17**	**66**	**21**	**54**	**136**	**14**	**21**	**77**	**26**	**32**	**19**	**41**	**6**	**47**	**45**	**141**	**193**	**18**	**6**	**40**	**35**	**1389**	
Producer's Accuracy	52.5	68.1	64.7	63.6	23.8	16.7	22.1	57.1	33.3	39.0	50.0	46.9	31.6	19.5	100.0	19.1	33.3	46.8	77.7	55.6	83.3	62.5	17.1		
Overall Accuracy	49.75																								

3E. CONFUSION MATRIX – MINOR PRUNING STRATEGY + SELECTED SIGNIFICANT BANDS

Minor Pruning + Selected Bands	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	47	1		3	1	2	4			3						3	2		2			2	2	72	65.3
AN		171		1	3	4	9		6	7				4	1		3		34			1	3	247	69.2
BD			12								2			4				8	1	1				28	42.9
CA	18			45	1	6	7	2		6			1			1	1					5	6	99	45.5
CC	3	3		2	7	1		1		2		2	3	2		3	2	4	4					39	17.9
CH	2	2				21	2			1							1						1	30	70.0
CI	5	5				2	33			1			2					17	3					68	48.5
CZ	1						1	12		1								1				1		17	70.6
DalMel				1					8	1						1						2	1	14	57.1
DC	22	3		11	3	12	12			46		5	1			4	10		1			1	6	137	33.6
DM			1				1				16							11	2					31	51.6
EN	1			2						1		18				3								25	72.0
GS	1			1	1	2	2			3			5			2	1							18	27.8
LC	3	7				1	1			1	2	1	1	8				5				1		31	25.8
LD								1	1				1		6	1		2						12	50.0
PA	1					2	7			1	1		3			7	1		1			1	1	26	26.9
PR	6	2		2	2	5	2		1	1		2				3	15		2			1		44	34.1
SA		1	7				1				2							65	3	2				81	80.2
SB		38		1			11			1	1			5		1		10	203	2		1		274	74.1
SchotB																		2		13				15	86.7
Stry																		2		3	4			9	44.4
TS	2				1	4	2	1		7		1	1				1		1			31	2	54	57.4
ZM	5	2		1		2	5	1	3	4		1	1			2	1	2	2			3	8	43	18.6
Grand Total	**117**	**235**	**20**	**70**	**19**	**64**	**100**	**18**	**19**	**87**	**24**	**30**	**19**	**23**	**7**	**31**	**38**	**129**	**259**	**21**	**4**	**50**	**30**	**1414**	
Producer's Accuracy	40.2	72.8	60.0	64.3	36.8	32.8	33.0	66.7	42.1	52.9	66.7	60.0	26.3	34.8	85.7	22.6	39.5	50.4	78.4	61.9	100.0	62.0	26.7		
Overall Accuracy	56.65																								

3F. CONFUSION MATRIX – TRADITIONAL MULTIPLE ENDMEMBER SAM (NO DECISION TREE)

Traditional Multiple Endmember SAM	AG	AN	BD	CA	CC	CH	CI	CZ	DalMel	DC	DM	EN	GS	LC	LD	PA	PR	SA	SB	SchotB	Stry	TS	ZM	Grand Total	User's Accuracy
AG	42	7		2	1	1	2			5	1			1		2	1	3	6					74	56.8
AN	9	172		1	2		6			8				2			4		35				10	249	69.1
BD			19								1	1		1				4	1	1				28	67.9
CA	4	1		58	3	3	1			9			1					1	11			5	3	100	58.0
CC	1	2		2	15	1		1		1				3			3	9	1					39	38.5
CH	1	4	1			19	2			1							1						1	30	63.3
CI	6	4		2	1	2	35	1		4		1	1	2				4	5					68	51.5
CZ	1						1	9		1								2	2			1		17	52.9
DalMel		4		1					6							1						1	1	14	42.9
DC	7	11	1	8	3	3	8			65		4		2			6	2	17			1	4	142	45.8
DM	1						1				23	1				2		1	2					31	74.2
EN	1	1		2						1		16				1		3						25	64.0
GS	1	3		1		2	1			1			5	1		1		2	1					19	26.3
LC	6	5				2	1				1	1		11				3				1		31	35.5
LD								1	1				1		6			3						12	50.0
PA	3	2					6				2		1	1		7		2				1	1	26	26.9
PR	1	2		4	3	2	1			6	2	1	1			2	20		1				1	47	42.6
SA	1		4	5							2	6			1	3		53	3	1	1		1	81	65.4
SB	16	23		7	1	2	10	4		8	2	2		4		9		2	165			12	7	274	60.2
SchotB																		2		13				15	86.7
Stry																				2	6			8	75.0
TS		4			1	2				2		1		1			1	2	6			34		54	63.0
ZM	2	6		4		1	2	1	1	3	1						2	2	8			2	12	47	25.5
Grand Total	**103**	**251**	**25**	**97**	**30**	**40**	**77**	**17**	**8**	**115**	**35**	**34**	**10**	**29**	**7**	**28**	**38**	**100**	**264**	**17**	**7**	**58**	**41**	**1431**	
Producer's Accuracy	40.8	68.5	76.0	59.8	50.0	47.5	45.5	52.9	75.0	56.5	65.7	47.1	50.0	37.9	85.7	25.0	52.6	53.0	62.5	76.5	85.7	58.6	29.3		
Overall Accuracy	56.67																								